SELECTED TOPICS IN MODERN ALGEBRA

by

mil

E. ARTIN

A Summer Conference in Collegiate Mathematics
Sponsored by the National Science Foundation
The University of North Carolina, 1954

Notes by J. H. Wahab

Artin, SELECTED TOPICS IN MODERN ALGEBRA

ERRATA

Change

page	line		to
1	- 6	a union	the union
9	3	is	are
9	6	are	is
13	12	chain	$\underline{\text{chain}}$
14	11	a $\underline{\text{composition}}$ series	a $\underline{\text{composition series}}$
15	- 6	projective	projective groups
18	15	$\underline{\ p\ }$	$\xrightarrow{p}$
40	- 8	dependent	independent
43	15	$Q[x]$	$Q(x)$
68	2	their	its
71	2	$\sigma(\tau(x)$	$\sigma(\tau(x))$
78	3	$E = (\alpha_1, \alpha_2, \ldots, \alpha_n)$	$E = F(\alpha_1, \alpha_2, \ldots, \alpha_n)$
86	-12	ξ	ζ
89	1	$\prod_{i=0}^{i=\alpha+p-1}$	$\prod_{i=0}^{p-1}$
89	- 5	another	also a

CHAPTER I

SETS AND MAPS

1.1 Notation and operations. The concept of a set of elements is basic in mathematics. Examples of sets of numbers and symbols reserved for them follow: Z, the set of integers; Q, the set of rational numbers; R, the set of real numbers; C, the set of complex numbers. The statement "3 is an element of Z" is abbreviated $3 \in Z$. The symbol $\notin$ is used to indicate that an element does not belong to a particular set, e.g. $2/3 \notin Z$. If every element of a set A is also an element of a set B, A is called a subset of B and B is said to contain A; this is written $A \subset B$ or $B \supset A$. The equality $A = B$ means $A \subset B$ and $B \subset A$. A set may be defined by enumerating its elements or by giving characteristic properties of its elements. The set of positive even integers less than 10 could be represented by $\{2, 4, 6, 8\}$, read "the set whose elements are 2, 4, 6, 8," or by $\{ x \mid x \in Z, x \text{ is even}, 0 < x < 10 \}$, read "the set of all x such that x is an integer, x is even, and x is greater than zero and less than 10." This set may also be described by $\{ x \}$ where $x \in Z$, x is even, and $0 < x < 10$.

The set of all elements belonging to either a set A or a set B is called ~~a~~ the union, $A \cup B$, of A and B. Those elements common to both A and B comprise a set known as the intersection, $A \cap B$, of A and B. If $A \subset B$, the complement of A relative to B is defined as the elements of B not in A. The empty set enables the intersection of sets to be closed and is the complement of a set A relative to A. The symbol $\emptyset$ will be reserved for the empty set. If to each element α of a set I there corresponds a

set A_α, then the collection (set) $\{ A_\alpha \}$ of all such A_α forms an indexed family of sets with the index α ranging over I. The definitions of union and intersection are extended to an indexed family of sets as follows:

$$\bigcup_\alpha A_\alpha = \{ x \mid x \in A_\alpha \text{ for at least one } \alpha \}$$

$$\bigcap_\alpha A_\alpha = \{ x \mid x \in A_\alpha \text{ for all } \alpha \} .$$

The cartesian product, $A \times B$, of two sets A and B is the set of all (ordered) pairs (a,b), a in A and b in B. If I is the set composed of the integers 1 and 2, each pair (a,b) of $A \times B$ can be thought of as a map of 1 into A and 2 into B. Hence, the cartesian product is the collection of all maps of I into $A \cup B$ such that 1 maps onto an element of A and 2 maps onto an element of B. (A definition of a mapping will appear later.) In general, if $\{ A_\alpha \}$ is an indexed family of sets, the cartesian product $\prod_\alpha A_\alpha$ of the sets in the family is the collection of all maps of I into $\bigcup_\alpha A_\alpha$ such that the image of α is an element of A_α. $A \times B \times C$ consists of all triples (a,b,c), a in A, b in B, c in C, and for A = B = C = R becomes the set R^3 of points in 3-dimensional space. In terms of the cartesian product of a family of sets, the axiom of choice can be stated simply as follows: If no A_α is empty, $\prod_\alpha A_\alpha$ is not empty.

1.2 Equivalence Relations. Let $\mathcal{R}$ be a subset of the set of pairs (a,b) comprising $A \times A$, a in A, b in A. If $(a,b) \in \mathcal{R}$, a is said to have the binary relation $\mathcal{R}$ to b, written $a \sim b$. This binary relation $\mathcal{R}$ over the set A is defined to be an equivalence relation if it has the further properties:

R (reflexive): $a \sim a$
S (symmetric): $a \sim b \Longrightarrow b \sim a$ ($\Longrightarrow$ is read "implies")
T (transitive): $a \sim b,\ b \sim c \Longrightarrow a \sim c$

If $\mathcal{R}$ is an equivalence relation over the set A and $S_a = \{ x \mid x \varepsilon A, x \sim a \}$, then

1) any $a \varepsilon A$ lies in one S_a,

2) $S_a \cap S_{a'} \neq \emptyset \Longleftrightarrow S_a = S_{a'} \Longleftrightarrow a \sim a'$.

The collection $\{S_a\}$ of sets splits A into subsets, called equivalence classes, which are disjoint in pairs. Any collection of subsets of a set A such that each element of A belongs to one and only one of the subsets is defined as a partition of the set A; the partition is called the factor or quotient set. Hence, an equivalence relation $\mathcal{R}$ over a set A induces a partition $A/\mathcal{R}$ (read "A modulo $\mathcal{R}$") of the set. Conversely, it can be shown that any partition of a set A may be used to define an equivalence relation over A by taking $a \sim a'$ to mean that a and a' belong to the same subset in the partition.

1.3 Mappings. A mapping f of a set A into a set B is a correspondence such that with each element a in A there is associated in B a single element f(a), called the image of a under f or the value of f at a. Such a mapping is written

$$f : A \longrightarrow B \quad \text{or} \quad A \xrightarrow{f} B .$$

The mapping f is sometimes called a function ranging over A with values in B. Equality of two mappings $f : A \longrightarrow B$ and $f' : A' \longrightarrow B'$ means $A = A'$, $B = B'$, and $f(a) = f'(a)$ for every a in A. If D is a subset of A, the set of images of the elements of D is designated by f(D). A mapping f of a set A onto a set B means further that $f(A) = B$. A mapping is said to be one-one provided $f(a) = f(a')$ implies $a = a'$, i.e. provided each element in f(A) is the image of only one element in A.

From two mappings $f : A \longrightarrow B$ and $g : B \longrightarrow D$ a composite map, called the product gf, of A into D can be constructed by associating with each element a in A the element $g[f(a)]$ in D. The composition of maps is associative, i.e. if $A \xrightarrow{f} B \xrightarrow{g} D \xrightarrow{h} E$, then $(hg)f = h(gf)$.

A mapping $i : A \longrightarrow B$ where $A \subset B$ and $i(a) = a$ for every a in A is called the injection map of A into B. The injection map is one-one; if $A = B$, it is also onto and is known as the identity map of A. From an equivalence relation $\mathcal{R}$ over a set A, a projection map p of A into $A/\mathcal{R}$ can be made by letting $p(a) = S_a$ where S_a is the equivalence class to which a belongs. An arbitrary map $f : A \longrightarrow B$ can be decomposed into the product of three canonical maps by defining the equivalence relation $\mathcal{R}$ natural to associate with f, i.e.

$$a_1 \sim a_2 \Longleftrightarrow f(a_1) = f(a_2),$$

and introducing the maps

$$p : A \longrightarrow A/\mathcal{R} \quad \text{where } p(a) = S_a ,$$

$$g : A/\mathcal{R} \longrightarrow f(A) \quad \text{where } g(S_a) = f(a)$$

$$(g \text{ is well-defined}),$$

$$i : f(A) \longrightarrow B \quad \text{where } i \text{ is the injection map.}$$

It follows that $f = igp$. Note that p is onto, i is one-one, and g is one-one onto.

Example: If $A = B = R$ and $f(a) = a^2$, then the decomposition of f is given by

$$R \xrightarrow{p} R/\mathcal{R} \xrightarrow{g} R^+ \xrightarrow{i} R$$

where R^+ is the set of non-negative real numbers and $a_1 \sim a_2$ means $a_1{}^2 = a_2{}^2$. Note that $S_3 = \{3, -3\} = S_{-3}$.

In the mapping $f : A \rightarrow B$, the elements of A that map into a subset D of B are designated by $f^{-1}(D)$. Note that $f[f^{-1}(D)]$ may not contain all of the elements of D and $f^{-1}[f(E)]$, where E is a subset of A, may contain elements not in E. However, if f is one-one onto, with every b in B may be associated the unique element a in A such that $f(a) = b$. This mapping, which is also one-one onto, is called the inverse mapping f^{-1} of B into A. Observe that $f^{-1}f$ and ff^{-1} are the identity maps on A and B respectively.

1.3 Partially ordered sets. A set A is partially ordered by the binary relation $<$ provided:

1) $a \leq b,\ b \leq a \Rightarrow a = b$,

2) $a < b,\ b < c \rightarrow a < c$.

An element a of a partially ordered set is called maximal if $a \leq x \Longrightarrow x = a$. A subset B of a partially ordered set A has an upper bound if there exists an a in A such that $b \leq a$ for every b in B. The set A of subsets of a given set are partially ordered by the relation $\subset$. This example shows that two elements of a partially ordered set need not be comparable; however, if one and only one of the statements, $a < b$, $a = b$, $(b < a)$ is true, the set is said to be totally ordered. A partially ordered set A is inductively ordered if every totally ordered subset of A has an upper bound (in A).

Zorn's Lemma (a set-theoretical transfinite axiom equivalent to the axiom of choice): An inductively ordered set contains at least one maximal element.

CHAPTER II

GROUPS, RINGS, AND FIELDS

2.1 Groups. Let G be a non-empty set and f a mapping of $G \times G$ into G. $ab = c$ will be used to indicate that f maps the pair (a, b) onto c. The set G is called a group if the following hold:

1) Associative law: $(ab)c = a(bc)$,
2) Existence of left unit: There exists at least one element e in G such that for every a in G $ea = a$,
3) Existence of left inverse: For every a in G there corresponds an element a^{-1} in G such that $a^{-1}a = e$.

Should in 2) or 3) a multiple choice be available, it is assumed that a definite choice has been made.

Exercise A: Show that $a_1 a_2 a_3 \ldots a_n$ is meaningful, i.e. generalize the associative law.

Exercise B: Show that a group is not necessarily obtained if the existence of a right inverse is substituted for 3) above.

Since $(a^{-1})^{-1}\left[(a^{-1}a)a^{-1}\right] = \left[(a^{-1})^{-1}a^{-1}\right]\left[aa^{-1}\right]$, $e = aa^{-1}$, and a left inverse is a right inverse. From $a(a^{-1}a) = (aa^{-1})a$ it follows that a left unit is also a right unit. It is easily shown that the equation $a\,x\,b = c$ has the unique solution $x = a^{-1}\,c\,b^{-1}$. Hence, e and a^{-1}, being solutions of $xa = a$ and $xa = e$, respectively, are unique. Moreover, $(a^{-1})^{-1} = a$ as each is a solution of $xa^{-1} = e$.

Clearly, $(a_1a_2\dots a_n)^{-1} = a_n^{-1}\dots a_2^{-1}a_1^{-1}$. From the definitions, $a^n = aa\dots a$ (n factors), $a^0 = e$ (or 1), $a^{-n} = (a^{-1})^n$, it follows that $a^n \cdot a^m = a^{n+m}$ and $(a^m)^n = a^{mn}$ where m and n are integers. To signify that c is associated with the ordered pair (a,b), $a + b = c$ is sometimes used instead of $ab = c$, and the group is called additive or multiplicative accordingly. For obvious psychological reasons the notation is adjusted for additive groups, e.g. the unit element is written as 0, the inverse of a as $-a$, and the n-th power of a as na.

2.2 <u>Equivalence Relations over Groups; Subgroups</u>. The interesting equivalence relations over a group are naturally related to the group operation (multiplication). For instance, an equivalence relation $\sim$ over a group G such that $a \sim b \Rightarrow ca \sim cb$ is described as <u>left stable</u>. For a left stable equivalence relation, $a \sim b \Leftrightarrow b^{-1}a \sim 1$; consequently, the equivalence class to which a belongs can be defined in terms of the equivalence class containing 1, which shall be studied. Let H be the equivalence class of 1, i.e. $H = \{ a \mid a \varepsilon G, a \sim 1\}$; then $a \sim b \Leftrightarrow b^{-1}a \varepsilon H$. Further, $a \varepsilon H, b \varepsilon H \Rightarrow a \sim 1, b \sim 1 \Rightarrow a \sim 1, ab \sim a \Rightarrow ab \varepsilon H$, and $a \sim 1 \Rightarrow a^{-1}a \sim a^{-1} \Rightarrow a^{-1} \sim 1 \Rightarrow a^{-1} \varepsilon H$. Therefore H is a subset of G that is closed under the group operation and contains the inverse of each of its elements. This characterizes H as a <u>subgroup</u> of G, i.e. a subset of G that forms a group under the binary operation of G. Moreover, if S_b is the equivalence class of an element b in G, then

$$S_b = \{ a \mid a \sim b\} = \{ a \mid b^{-1}a \sim 1\} = \{ a \mid b^{-1}a \varepsilon H\} = \{ bh \mid h \varepsilon H\}$$

where the last set is abbreviated bH and is called a <u>loft coset</u> of H in G.

Conversely, if H is a subgroup of G, it will be shown that the binary relation $a \sim b$ meaning $b^{-1}a \varepsilon H$ is a left stable equivalence relation over G having H as the equivalence class containing 1. First, this relation is an equivalence relation since

$$R : a^{-1}a = 1 \varepsilon H \Rightarrow a \sim a ,$$

$$S : a \sim b \Rightarrow b^{-1}a \varepsilon H \Rightarrow (b^{-1}a)^{-1} \varepsilon H \Rightarrow a^{-1}b \varepsilon H \Rightarrow b \sim a,$$

$$T : a \sim b,\ b \sim c \Rightarrow b^{-1}a \varepsilon H,\ c^{-1}b \varepsilon H \Rightarrow c^{-1}bb^{-1}a \varepsilon H$$
$$\Rightarrow c^{-1}a \varepsilon H \Rightarrow a \sim c.$$

Secondly, it is left stable because $a \sim b \Rightarrow b^{-1}a \varepsilon H \Rightarrow (cb)^{-1}ca \varepsilon H \Rightarrow ca \sim cb$. Finally, as $a \sim 1 \Leftrightarrow 1^{-1}a = a \varepsilon H$, H is the equivalence class containing 1. Consequently, the set of left stable equivalence relations over a group G is in a one-one correspondence with the set of subgroups of G. The partition of G induced by the loft stable equivalence relation is the family {bH} of left cosets of H in G, i.e. $G = \bigcup_b bH$ and either $bH \cap cH = \emptyset$ or $bH = cH$. Similar results can be obtained for a <u>right stable</u> equivalence relation ($a \sim b \Rightarrow ac \sim bc$).

A <u>stable</u> equivalence relation is one that is both left stable and right stable ($a \sim b \Rightarrow ca \sim cb$ and $ac \sim bc$). For H as defined above, left stability implies that the equivalence class S_b of an element b in G is the left coset bH, but right stability similarly implies that $S_b = Hb$ (note that H was defined independently of left stability). Hence, for a stable equivalence relation every left coset is a right coset, $bH = Hb$, which is equivalent to $bHb^{-1} = H$ for every b in G, the defining property of an <u>invariant</u> (<u>normal</u>) subgroup H in G. This definition of an invariant subgroup H in G may be weakened to $bHb^{-1} \subset H$ for every b in G since

$b^{-1}Hb \subset H$ and therefore $H = bb^{-1}Hbb^{-1} \subset bHb^{-1}$. For stable equivalence relations over G it is important to note that only stability and the closure of the group multiplication are ~~is~~ needed to prove: $a \sim b,\ c \sim d \Rightarrow ac \sim bd$ (the intermediate step is $ac \sim bc$ and $bc \sim bd$). The existence of an inverse in G gives further that $a \sim b \Rightarrow a^{-1} \sim b^{-1}$. The set of stable equivalence relations over G is ~~are~~ in a one-one correspondence with the set of invariant subgroups of G.

When an algebraic system such as a group is partitioned into equivalence classes by a stable equivalence relation, it is customary to attempt to form a new algebraic structure whose elements are the equivalence classes. The general procedure will be followed for a group. To multiply two equivalence classes, a representative element from each is selected, these elements are multiplied in the proper order to obtain a new element, and the product of the equivalence classes is the equivalence class to which this new element belongs. Needed to show that this product is well-defined is precisely that $a \sim b,\ c \sim d \Rightarrow ac \sim bd$. The multiplication of the equivalence classes is associative; in the new structure the class containing 1 is the unit element, and the inverse of the class containing a is the class containing a^{-1}. Hence, the set of equivalence classes is a group, which is called the <u>factor</u> (<u>quotient</u>) <u>group</u>, G modulo H, written G/H. Note that the factor group is written in terms of G and H rather than in terms of G and the equivalence relation $\mathcal{R}$ corresponding to H. In the case of groups, a simplified definition of the multiplication of equivalence classes works, viz. $aH \cdot bH = \{ ah_1 \cdot bh_2 \mid h_1, h_2 \in H \}$. However, crucial to the argument that this is the same multiplication is the property $H \cdot H = H$.

2.3 Homomorphisms of Groups. A homomorphism of a group G into a group H is a map, $f : G \rightarrow H$, for which the group operation is preserved, i.e. $f(ab) = f(a) \cdot f(b)$ It will be assumed that the context makes clear whether the group operation in G or H is meant and whether 1 designates the unit of G or the unit of H. A homomorphism also preserves the unit element and inverses because $f(a) = f(1 \cdot a) = f(1) \cdot f(a)$ and $f(a^{-1}) \cdot f(a) = f(a^{-1}a) = f(1) = 1$.

Let K be an invariant subgroup of G (H had this role in Sec. 2.2) and let p be the canonical map (Sec. 1.3) of G onto G/K, i.e. $p : G \rightarrow G/K$ where $p(a) = aK$. Since $p(ab) = abK = aK \cdot bK = p(a) \cdot p(b)$, p is a homomorphism. Such a homomorphism will now be shown to be the only interesting factor in the decomposition of an arbitrary homomorphism $f : G \rightarrow H$ into the product igp according to Sec. 1.3 and in this sense may be thought of as the only homomorphism of a group. The natural equivalence relation associated with f, $a \sim b \Longleftrightarrow f(a) = f(b)$, is left stable for $f(a) = f(b) \Longrightarrow f(c) \cdot f(a) = f(c) \cdot f(b) \Longrightarrow f(ca) = f(cb) \Longrightarrow ca \sim cb$. Similarly, it is right stable, consequently stable, and therefore $\{ a \mid a \varepsilon G, a \sim 1 \}$ is the desired invariant subgroup K. As $a \sim 1 \Longleftrightarrow f(a) = f(1) = 1$, $K = \{ a \mid a \varepsilon G, f(a) = 1 \}$. This group K is defined as the kernel of the homomorphism f. Hence, $G \xrightarrow{p} G/K \xrightarrow{g} f(G) \xrightarrow{i} H$ and $f = igp$, where p has been shown to be a homomorphism and i is clearly a homomorphism. That g is also a homomorphism follows from $g(aK \cdot bK) = g(abK) = f(ab) = f(a) \cdot f(b) = g(aK) \cdot g(bK)$. g is both a homomorphism and one-one onto; such a mapping is called an isomorphism, and G/K is said to be isomorphic to f(G), written $G/K \simeq f(G)$.

Let K be any invariant subgroup of G and p be the canonical

homomorphism $p : G \rightarrow G/K$. The kernel of p is then K (taking f as p makes g and i merely identity mappings on G/K). Further, let U be a subgroup of G and q be the mapping p restricted to U, i.e. $q : U \rightarrow G/K$ where $q(a) = p(a) = aK$, $a \varepsilon U$ (q is a homomorphism). Consequently, $q(U) = \{ uK \mid u \varepsilon U \}$ and $p^{-1}\left[q(U)\right] = UK \supset U$. That K is invariant yields $UK \cdot UK = UUKK = UK$ and $(UK)^{-1}$, defined as $\{(uk)^{-1} \mid u \varepsilon U, k \varepsilon K\}$, equals $K^{-1}U^{-1} = KU = UK$. Hence, UK is a subgroup of G and has K as an invariant subgroup. As $uk \cdot K = uK$, $UK/K = q(U)$. The kernel of q is obviously $K \cap U$. It has been shown that the homomorphism $q : U \rightarrow G/K$ has $K \cap U$ as kernel and UK/K as the image of U; therefore, by the isomorphism of the preceding paragraph, $U/K \cap U \simeq UK/K$. This result is known as one of the isomorphism theorems of groups.

Exercise A: Generalize the results concerning groups to groups with operators.

Exercise B: Show that the intersection of two subgroups of a group G is a group.

2.4 Jordan-Hölder Theorem. Let G be a group having four subgroups U, U_o, V, V_o where U_o is an invariant subgroup of U and V_o is an invariant subgroup of V.

Lemma: $(U \cap V_o)U_o$ is an invariant subgroup of $(U \cap V)U_o$.

Proof: U_o is an invariant subgroup of U and $U \cap V$ is a subgroup of U. Hence, the results of the last paragraph of Sec. 2.3 may be applied. $(U \cap V) \cdot U_o$ (corresponding to UK in the paragraph mentioned) is a subgroup of U; and, since U_o is an invariant subgroup of U, $(U \cap V)U_o = U_o(U \cap V)$. Similarly, $U \cap V_o$ is a subgroup of U and $(U \cap V_o)U_o = U_o(U \cap V_o)$, a subgroup

of U. Clearly, $(U \cap V_o)U_o \subset (U \cap V)U_o$ and the "subgroup part" of the lemma is established. Further, by the isomorphism theorem of Sec. 2.3,

$$U \cap V/U_o \cap (U \cap V) \simeq (U \cap V)U_o/U_o ,$$

and in particular $U_o \cap V$ [$= U_o \cap (U \cap V)$] is an invariant subgroup of $U \cap V$. From the symmetrical nature of the hypothesis, $V_o \cap U$ is also an invariant subgroup of $V \cap U$. The tools are now available to complete the proof by showing that ab, where a ranges over $U \cap V$ and b ranges over U_o [ab then ranges over $(U \cap V)U_o$], commutes with $(U \cap V_o)U_o$. This will be accomplished by showing that each of a and b commutes separately.

1) $a(U \cap V_o)U_o = (U \cap V_o)aU_o = (U \cap V_o)U_o a$, since $a \varepsilon U \cap V$ which has $U \cap V_o$ as an invariant subgroup and $a \varepsilon U$ which has U_o as an invariant subgroup,

2) $b(U \cap V_o)U_o = bU_o(U \cap V_o) = U_o(U \cap V_o) = (U \cap V_o)U_o = (U \cap V_o)U_o b$, where $b \varepsilon U_o$ is needed for the second and fourth equalities.

Whereas a is "pulled through," b is carried through by U_o.

<u>Main lemma</u>: With the four subgroups as in the lemma,

$$(U \cap V)U_o/(U \cap V_o)U_o \simeq (U \cap V)V_o/(U_o \cap V)V_o .$$

Proof: Notice that because of the lemma each of the above members makes sense. The elements of the factor group on the left are simply the cosets of $(U \cap V_o)U_o$ in $(U \cap V)U_o$ and, with a and b as in the proof of the lemma, are given by $ab(U \cap V_o)U_o$ or $a(U \cap V_o)U_o$ since from 2) above the b may be absorbed. Similarly, the elements of the factor group on the right are $a(U_o \cap V)V_o$. The desired isomorphism is given by the obvious mapping, viz. $a(U \cap V_o)U_o \longleftrightarrow a(U_o \cap V)V_o$. Demonstrating $a_1(U \cap V_o)U_o = a_2(U \cap V_o)U_o \Longleftrightarrow a_1(U_o \cap V)V_o = a_2(U_o \cap V)V_o$, where

$a_1, a_2 \varepsilon U \cap V$, will confirm that this mapping is well-defined. The following suffices: $a_1(U \cap V_0)U_0 = a_2(U \cap V_0)U_0 \Longleftrightarrow a_2^{-1}a_1 \varepsilon (U \cap V_0)U_0$ $\Longleftrightarrow a_2^{-1}a_1 = \alpha\beta$, $\alpha \varepsilon U \cap V_0$, $\beta \varepsilon U_0$, $\alpha\beta \varepsilon U \cap V \Longleftrightarrow a_2^{-1}a_1 = \alpha\beta$, $\alpha \varepsilon U \cap V_0$, $\beta \varepsilon U_0 \cap V$ [Note that $\beta \varepsilon V$ follows from $\alpha\beta \varepsilon V$, since $U \cap V \subset V$, and from $\alpha^{-1} \varepsilon V$, since $\alpha \varepsilon U \cap V_0$ and $U \cap V_0 \subset V$.] $\Longleftrightarrow$ $a_2^{-1}a_1 \varepsilon (U \cap V_0)(U_0 \cap V)$; by symmetry, $a_1(U_0 \cap V)V_0 = a_2(U_0 \cap V)V_0$ $\Longleftrightarrow a_2^{-1}a_1 \varepsilon (U_0 \cap V)(U \cap V_0)$; and $(U \cap V_0)(U_0 \cap V) = (U_0 \cap V)(U \cap V_0)$ since $U_0 \cap V$ is an invariant subgroup of $U \cap V$. Clearly, this map is one-one onto. That it is a homomorphism follows from the fact that $(U \cap V_0)U_0$ is an invariant subgroup of $(U \cap V)U_0$. Therefore, it is an isomorphism.

A chain for the group G is a set of subgroups G_i of G such that $G = G_r \supset G_{r-1} \supset \dots \supset G_0 = 1$ and G_{i-1} is an invariant subgroup of G_i for $i = 1, 2, \dots, r$. Let $G = H_s \supset H_{s-1} \supset \dots \supset H_0 = 1$ be a second chain for G, i.e. H_{k-1} is an invariant subgroup of H_k for $k = 1, 2, \dots, s$. The main lemma applied to the subgroups G_i, G_{i-1}, H_k, H_{k-1} (for U, U_0, V, V_0, respectively) gives

$$(G_i \cap H_k)G_{i-1}/(G_i \cap H_{k-1})G_{i-1} \simeq (G_i \cap H_k)H_{k-1}/(G_{i-1} \cap H_k)H_{k-1} ,$$

by virtue of which another chain for G can be constructed. Let each segment $G_i \supset G_{i-1}$ of the first chain be replaced by

$$G_i = (G_i \cap H_s)G_{i-1} \supset (G_i \cap H_{s-1})G_{i-1} \supset \dots \supset (G_i \cap H_0)G_{i-1} = G_{i-1} .$$

The resulting chain, since it contains every G_i, is called a <u>refinement</u> of the first chain. In a similar way, the first chain can be used to

construct a refinement of the second chain with each segment $H_k \supset H_{k-1}$ replaced by $H_k = (G_r \cap H_k)H_{k-1} \supset (G_{r-1} \cap H_k)H_{k-1} \supset \dots \supset (G_o \cap H_k)H_{k-1} = H_{k-1}$. If each factor group G_i/G_{i-1} is isomorphic to a factor group H_k/H_{k-1} and conversely, then the first and second chains are said to be isomorphic. Consequently, from the above isomorphism the two refinements are isomorphic, and a theorem may be stated.

Theorem of Schreior: Any two chains for a group have isomorphic refinements.

A refinement of a chain for a group is defined as proper if the refinement contains a group not in the original chain. A chain which has no proper refinement is called a composition series. Two composition series for a group have isomorphic refinements; however, as the refinements cannot be proper, all additional factor groups must be trivial. Hence, two composition series of the same group are isomorphic (Jordan-Hölder Theorem). Hölder noted that for finite groups the orders of the factor groups of one composition series will be the same as the orders of the factor groups in another.

Exercise: If G has one composition series, then any other chain can be refined to a composition series.

The definition of a proper refinement raises the following question: Given that K is an invariant subgroup of group G, does there exist a group G_o such that G_o contains K properly and is a proper invariant subgroup of G? The answer is no if and only if the group G/K is simple. A group G is called simple if the only invariant subgroups of G are the two trivial ones. To show this, consider a mapping of G onto a group H with kernel K: $f : G \longrightarrow H \ (\simeq G/K)$. For instance, f could be the canonical map of G

onto G/K. First, let H_o be a subgroup of H and $G_o = f^{-1}(H_o)$. G_o is a subgroup of G for $f(a) \varepsilon H_o$, $f(b) \varepsilon H_o \Longrightarrow f(ab) = f(a)f(b) \varepsilon H_o \Longrightarrow ab \varepsilon G_o$, and $f(a^{-1}) = [f(a)]^{-1} \varepsilon H_o \Longrightarrow a^{-1} \varepsilon G_o$; and $K \subset G_o$ since $f(K) = 1 \varepsilon H_o$. Conversely, if G_o is given, let $H_o = f(G_o)$, a subgroup of H (it is easy to show that the homomorphic image of a group is a group). Then form $f^{-1}(H_o) = G_1$. $G_o \subset G_1$ since $f(G_o) = H_o$. To show $G_1 \subset G_o$, let $a \varepsilon G_1$, i.e. $f(a) = h_o \varepsilon H_o$. Since $f(G_o) = H_o$, there exists $b \varepsilon G_o$ such that $f(b) = h_o$ and $f(b^{-1}) = h_o^{-1}$. Hence, $f(b^{-1}a) = f(b^{-1})f(a) = h_o^{-1}h_o = 1$, which implies $b^{-1}a \varepsilon K \subset G_o$ or $a \varepsilon bG_o = G_o$. Therefore, $G_1 = G_o$. Next, suppose G_o is an invariant subgroup of G and $H_o = f(G_o)$. Then, an element of H can be written f(a), $f(a)f(G_o)f(a^{-1}) = f(aG_oa^{-1}) = f(G_o) = H_o$, and H_o is invariant in H. Finally, if H_o is invariant in H, $f(aG_oa^{-1}) = f(a)f(G_o)f(a^{-1}) = f(G_o) = H_o$; therefore, $aG_oa^{-1} \subset G_o$ and G_o is invariant in G.

Which finite groups are simple is unknown. Of course, those of prime order are simple. Some simple groups of non-prime order are: the icosahedron group, order 60; projective group of the plane modulo 2, order 168; alternating group of 6 elements, order 360; and the projective groups on lines of orders 504, 660, 1092. Moreover, there is only one simple group for each of these orders. The smallest non-prime order for which there are two simple groups is 20160; however, there exist infinitely many such orders. For each prime $p \geq 5$, there exists a simple group of order $\frac{1}{2} p(p^2 - 1)$.

2.5 The Additive Group of Integers, Z. It will first be noted that a group G of order N is partitioned by a subgroup H of order n into disjoint left cosets. Let the number of cosets be j, called the index of H in G. Since any two cosets contain the same number of elements, $N = nj$. The set Z under addition forms a commutative or abelian ($ab = ba$) group, called the additive group of integers. A study will be made of the subgroups H of Z. Certainly, $0 \varepsilon H$, and the set $\{ 0 \}$ consisting of 0 alone is a subgroup. If, however, $H \neq \{ 0 \}$, H must contain a smallest positive integer, d, and $H = dZ$ as follows:

1) $dZ \subset H$ since $md = d + d + \ldots + d, 0,$ or $(-d) + (-d) + \ldots + (-d)$, each of which belongs to H;

2) $H \subset dZ$ since, by long division, $a \varepsilon H \Longrightarrow a = qd + r, 0 \leq r < d \Longrightarrow r = (a - qd) \varepsilon H, 0 \leq r < d \Longrightarrow r = 0$, otherwise d would not be the smallest positive integer in H. Of course, these subgroups are invariant in Z (commutative).

The additive group of integers, Z, can be used to analyze the structure of an arbitrary group G since, corresponding to each element a in G, there is a homomorphism, $f : Z \longrightarrow G$, given by $f(n) = a^n$. $f(n)f(m) = a^n \cdot a^m = a^{n+m} = f(n + m)$, f(Z) is the set of all powers of a, and the kernel K of f is dZ where $d = 0$, or d is a positive integer. Moreover, $Z/dZ \simeq f(Z)$. For $d = 0$, this becomes $Z/\{0\} \simeq f(Z)$, hence $Z \simeq f(Z)$; this is the case when no two powers of a are equal. For $d > 0$, $a^d = 1$, and d is called the period or order of the element a, as d is the smallest positive exponent for which the corresponding power of a is 1. $1, a, a^2, \ldots, a^{d-1}$ are distinct and $\{1, a, a^2, \ldots, a^{d-1}\} = f(Z)$. Since d is the order of the subgroup f(Z) of G, the order of an element in a group must divide the order of the group.

2.6 <u>Rings</u>. A non-empty set $\mathscr{R}$ with two binary operations, + and $\cdot$, is called a ring provided:

I. the set is a commutative group under addition (+),

II. the two distributive laws hold, i.e. $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$.

If the requirement that the addition be commutative is deleted, II applied to $(a + b)(c + d)$ yields $ac + bc + ad + bd = ac + ad + bc + bd$, whence $bc + ad = ad + bc$; i.e. commutativity of addition holds for those elements expressible as products of two elements even if it is not postulated in general. Hence, the commutativity of addition is implied if each element is expressible as a product, e.g. if the set contains a multiplicative unit. For an example in which the commutativity of addition is not implied, form a ring from a non-commutative group under addition by defining the product of any two elements to be 0.

If $\mathscr{R}$ is a ring, it follows that:

1) $0a = a0 = 0$,

2) $a(-b) = (-a)b = -(ab)$,

3) $(-a)(-b) = ab$,

since respectively

1) $0 = -(a0) + (a0) = -(a0) + (a[0 + 0]) = -(a0) + (a0) + (a0)$ $= 0 + a0 = a0$ (similarly $0 = 0a$),

2) $0 = a0 = a[b + (-b)] = ab + a(-b) \Longrightarrow a(-b) = -(ab)$ and $0 = 0b = [a + (-a)]b = ab + (-a)b \Longrightarrow (-a)b = -(ab)$,

3) replacing a by $-a$ in 2) gives $(-a)(-b) = [-(-a)]b = ab$.

Specialized rings are often studied by requiring besides I and II certain additional properties. Some of these properties and the corresponding nomenclature will be listed:

III. <u>Associative ring</u>, $a(bc) = (ab)c$;

IV. <u>Commutative ring</u>, $ab = ba$;

V. <u>Ring with a unit</u>, there exists an element $e \varepsilon \mathcal{R}$ such that $ea = ae = a$;

VI. <u>Domain of Integrity</u>, $ab = 0 \Longrightarrow a = 0$ or $b = 0$ (no proper divisors of zero);

VII. <u>Field</u>, non-zero elements form a group under multiplication;

VIII. <u>Lie Ring</u>, $a(bc) + b(ca) + c(ab) = 0$ (Jacobi identity) and $ab = - ba$.

2.7 <u>Equivalence Relations over Rings; Residue Class Rings</u>. Since a ring $\mathcal{R}$ forms a commutative group under addition, each additive subgroup $\mathcal{N}$ is an invariant subgroup. Consequently, from Sec. 2.2, the corresponding equivalence relation over $\mathcal{R}$, $c \sim d \Longleftrightarrow c - d \varepsilon \mathcal{N}$, is stable under addition. Moreover, $\mathcal{N} = \{a \mid a \varepsilon \mathcal{R}, a \sim 0\}$, the factor group is $\mathcal{R}/\mathcal{N}$ having as cosets $b + \mathcal{N}$, and the canonical map $\mathcal{R} \xrightarrow{p} \mathcal{R}/\mathcal{N}$ has $\mathcal{N}$ as kernel. If this equivalence relation is left stable under multiplication, then $a \sim b \Longrightarrow ca \sim cb$ for every $c \varepsilon \mathcal{R}$; for $b = 0$, this means $a \varepsilon \mathcal{N} \Longrightarrow ca \varepsilon \mathcal{N}$ for every $c \varepsilon \mathcal{R}$. Conversely, if $\mathcal{N}$ is an additive subgroup of $\mathcal{R}$ for which $a \varepsilon \mathcal{N} \Longrightarrow ca \varepsilon \mathcal{N}$ for every $c \varepsilon \mathcal{R}$, then $a \sim b \Longrightarrow a - b \varepsilon \mathcal{N} \Longrightarrow c(a - b) \varepsilon \mathcal{N} \Longrightarrow ca \sim cb$ and the equivalence relation corresponding to $\mathcal{N}$ is left stable under multiplication. An additive subgroup $\mathcal{N}$ of the ring $\mathcal{R}$ is defined as a <u>left ideal</u> of the ring if $a \varepsilon \mathcal{N} \Longrightarrow ca \varepsilon \mathcal{N}$ for every $c \varepsilon \mathcal{R}$. Hence, there is a one-one correspondence between the left ideals of a ring and the equivalence relations over the ring that are stable under addition and left stable under multiplication. Similar results hold for a right ideal (additive subgroup $\mathcal{N}$ such that $a \varepsilon \mathcal{N} \Longrightarrow ac \varepsilon \mathcal{N}$ for

every c ε $\mathfrak{r}$). If $\mathfrak{n}$ is both a left ideal and a right ideal, it is called simply an <u>ideal</u>. The set of ideals of a ring is in one-one correspondence with the set of equivalence relations over the ring that are stable under both addition and multiplication. If $\mathfrak{n}$ is an ideal, the factor group $\mathfrak{r}/\mathfrak{n}$ becomes a ring, called the <u>residue class ring</u> of $\mathfrak{r}$ modulo $\mathfrak{n}$, with the product of two residue (equivalence) classes defined in terms of the product of two representative elements, i.e. $(b + \mathfrak{n}) \cdot (c + \mathfrak{n}) = bc + \mathfrak{n}$. Only closure and stability are needed to show that this product is well-defined (Sec. 2.2). The distributive laws for the elements obviously yield the distributive laws for the residue classes. The associativity and commutativity, if they exist in the original ring, are likewise carried over to the residue class ring. If $\mathfrak{n}$ is an ideal, $a \sim b$, meaning $a - b \;\varepsilon\; \mathfrak{n}$, is written $a \equiv b \pmod{\mathfrak{n}}$, which is read a is congruent to b modulo $\mathfrak{n}$. Note that the simplified formulation of the product of cosets fails here as $\mathfrak{n}\,\mathfrak{n}$ is not necessarily $\mathfrak{n}$.

<u>Example</u>: Let $\mathfrak{r} = Z$. The additive subgroups are dZ and every subgroup is an ideal since multiplication in Z may be expressed in terms of addition. For the ring Z, $a \sim b$ is written $a \equiv b \pmod{d}$, the dZ being shortened to d.

<u>Convention</u>: Henceforth the rings studied, and consequently their residue class rings, will be assumed to be associative and commutative.

A ring is not necessarily a field; however, if it is an integral domain (domain of integrity), it may be embedded in its field of quotients, whose elements are the equivalence classes of the set of pairs a/b, $b \neq 0$, corresponding to the equivalence relation $a/b \sim c/d \Longleftrightarrow ad = bc$. The integral domain Z is embedded in the field Q in exactly this fashion.

It would, of course, be impossible to embed in a field a ring with divisors of zero. However, it may happen that the residue class ring modulo a certain ideal has no divisors of zero; in this event the ideal is called <u>prime</u> and denoted by $\mathfrak{p}$. Hence, $\mathfrak{p}$ is a prime ideal if and only if $ab \varepsilon \mathfrak{p}$, $a \notin \mathfrak{p} \Longrightarrow b \varepsilon \mathfrak{p}$, or $ab \equiv 0$, $a \not\equiv 0 \Longrightarrow b \equiv 0$ all mod $\mathfrak{p}$. If $\mathfrak{o} = Z$, $d \neq 0$, and $\mathfrak{p} = dZ$, this means $d \mid ab$, $d \nmid a \Longrightarrow d \mid b$, which is true from elementary number theory if and only if d is a prime number—the motivation for the name "prime" ideal.

Let $\mathfrak{o}$ be a ring with a unit element. An ideal $\mathfrak{m}$ in $\mathfrak{o}$ is called <u>maximal</u> if $\mathfrak{m} \neq \mathfrak{o}$ and if $\mathfrak{o}$ is the only ideal strictly above $\mathfrak{m}$. It will now be shown that $\mathfrak{m}$ is maximal if and only if $\mathfrak{o}/\mathfrak{m}$ is a field. First, assume $\mathfrak{m}$ a maximal ideal and let $a \notin \mathfrak{m}$. Consider the set $\mathfrak{b}$ of all elements of $\mathfrak{o}$ which have the form $\alpha + a\beta$ where α ranges over all elements of $\mathfrak{m}$ and β over those of $\mathfrak{o}$. Since $(\alpha_1 + a\beta_1) \pm (\alpha_2 + a\beta_2) = (\alpha_1 \pm \alpha_2) + a(\beta_1 \pm \beta_2)$ and $c(\alpha + a\beta) = c\alpha + a \cdot c\beta$, $\mathfrak{b}$ is an ideal. $\alpha = \alpha + a \cdot 0 \Longrightarrow \mathfrak{m} \subset \mathfrak{b}$, and $a \notin \mathfrak{m}$, $a = 0 + a \cdot 1 \varepsilon \mathfrak{b} \Longrightarrow \mathfrak{m} \neq \mathfrak{b}$. Therefore, $\mathfrak{b} = \mathfrak{o}$ as $\mathfrak{m}$ is maximal. Consequently, $1 \varepsilon \mathfrak{b}$, i.e. there exists an α in $\mathfrak{m}$ and a β in $\mathfrak{o}$ such that $1 = \alpha + a\beta$; this implies that $1 - a\beta \varepsilon \mathfrak{m}$, and $1 \equiv a\beta \pmod{\mathfrak{m}}$. Hence, residue class of 1 = residue class of a • residue class of β. The residue class of β is the inverse of the residue class of a since the residue class of 1 is clearly the multiplicative unit in $\mathfrak{o}/\mathfrak{m}$. Further, the residue class of a where $a \notin \mathfrak{m}$ is different from the zero residue class. It follows that $\mathfrak{o}/\mathfrak{m}$ is a field. Since a field has no divisors of zero, a by-product of this part of the proof is that every maximal ideal is prime. To prove the converse, assume $\mathfrak{o}/\mathfrak{m}$ to be a field and let $\mathfrak{b}$ be an ideal in $\mathfrak{o}$,

$\mathfrak{m} \subset \mathfrak{b}$, $\mathfrak{m} \neq \mathfrak{b}$. There exists in $\mathfrak{b}$ an element a not in $\mathfrak{m}$, and the residue class of a is not the zero residue class. Consequently, as $\mathfrak{o}/\mathfrak{m}$ is a field, the residue class of a has an inverse, i.e. there is a β in $\mathfrak{o}$ such that 1 = residue class of a $\cdot$ residue class of β. Whence, $a\beta \equiv 1$ (mod $\mathfrak{m}$), and $1 - a\beta \in \mathfrak{m}$. Since $\mathfrak{m} \subset \mathfrak{b}$, $1 - a\beta \in \mathfrak{b}$; also $a\beta \in \mathfrak{b}$ (a in the ideal $\mathfrak{b}$). Hence, $1 \in \mathfrak{b}$, which implies $\mathfrak{o} \subset \mathfrak{b}$. Therefore, $\mathfrak{o} = \mathfrak{b}$ and $\mathfrak{m}$ is maximal.

Let $\mathfrak{o}$ be a ring with unit element and $\mathfrak{m}$ be an ideal of $\mathfrak{o}$, $\mathfrak{m} \neq \mathfrak{o}$. Then, there exists a maximal ideal $\mathfrak{y}$ such that $\mathfrak{m} \subset \mathfrak{y}$. The existence of $\mathfrak{y}$ is given by Zorn's lemma as follows. Consider the set of all ideals $\mathfrak{t}$ where $\mathfrak{t} \supset \mathfrak{m}$, $\mathfrak{t} \neq \mathfrak{o}$. This set is partially ordered by inclusion. To show this set to be inductively ordered, suppose a totally ordered subset of $\{\mathfrak{t}\}$ to be represented by the indexed family $\{\mathfrak{t}_\alpha\}$. Construct $\mathfrak{d} = \bigcup_\alpha \mathfrak{t}_\alpha$. Clearly, $\mathfrak{d} \supset \mathfrak{t}_\alpha$ for each α and $\mathfrak{d} \supset \mathfrak{m}$. Needed, however, is that $\mathfrak{d}$ is an ideal and $\mathfrak{d} \neq \mathfrak{o}$. First, $\mathfrak{d}$ is an additive group: for $d_1, d_2 \in \mathfrak{d} \Longrightarrow d_1 \in \mathfrak{t}_{\alpha_1}$, $d_2 \in \mathfrak{t}_{\alpha_2}$; say $\mathfrak{t}_{\alpha_2} \subset \mathfrak{t}_{\alpha_1}$ ($\{\mathfrak{t}_\alpha\}$ totally ordered); hence, $d_1, d_2 \in \mathfrak{t}_{\alpha_1}$, which $\Longrightarrow d_1 \pm d_2 \in \mathfrak{t}_{\alpha_1}$ $\Longrightarrow d_1 \pm d_2 \in \mathfrak{d}$. Moreover, $d \in \mathfrak{d} \Longrightarrow d \in \mathfrak{t}_\alpha$ for some $\alpha \Longrightarrow \beta d \in \mathfrak{t}_\alpha$ $\Longrightarrow \beta d \in \mathfrak{d}$. Finally, $\mathfrak{d} \neq \mathfrak{o}$ for $1 \in \mathfrak{d} \Longrightarrow 1 \in \mathfrak{t}_\alpha$ for some $\alpha \Longrightarrow \mathfrak{t}_\alpha = \mathfrak{o}$, which is false.

Let $S = \{S_\alpha\}$ be a subset of a ring $\mathfrak{o}$ with unit element. The smallest ideal containing S is called the <u>ideal generated by S</u>. Clearly, $c_1s_1 + c_2s_2 + \ldots + c_ms_m$, where $c_i \in \mathfrak{o}$ and $s_i \in S$, must be an element of any ideal containing S and, in particular, the ideal generated by S. Moreover, the set $\mathfrak{a}$ of all elements of this type is an ideal, for $(c_1s_1 + c_2s_2 + \ldots + c_ms_m) \pm (c_1's_1' + c_2's_2' + \ldots + c_n's_n')$ and $d(c_1s_1 + c_2s_2 + \ldots + c_ms_m)$ where $d \in \mathfrak{o}$ are of this same type. $1 \cdot s \in \mathfrak{a}$, i.e. $S \subset \mathfrak{a}$. Therefore, $\mathfrak{a}$ is the ideal generated by S. Two special cases of ideals generated by a set follow:

1) Suppose $d \in \mathfrak{o}$ and $S = \{d\}$. Since $c_1d + c_2d + \ldots + c_md = (c_1 + c_2 + \ldots + c_m)d = cd$ where c_i and c are in $\mathfrak{o}$, $\mathfrak{a} = d \cdot \mathfrak{o}$. An ideal generated by a single element is called a <u>principal ideal</u>. Observe that for $\mathfrak{o} = Z$ every ideal is principal.

2) Given two ideals $\mathfrak{a}$ and $\mathfrak{b}$. It is easy to show that $\mathfrak{a} \cap \mathfrak{b}$ is an ideal, but $\mathfrak{a} \cup \mathfrak{b}$ need not be an ideal. However, the ideal generated by the union of $\mathfrak{a}$ and $\mathfrak{b}$ has some of the set properties of $\mathfrak{a} \cup \mathfrak{b}$. This ideal is the set of all elements $c_1a_1 + c_2a_2 + \ldots + c_ra_r + c_1'b_1 + c_2'b_2 + \ldots + c_s'b_s$ where $c_i, c_i' \in \mathfrak{o}$, $a_i \in \mathfrak{a}$, and $b_i \in \mathfrak{b}$. Obviously, each element in the generated ideal can be written as $a + b$ where $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ (a or b may be zero), and, conversely, each of the elements $a + b$ appears in the generated ideal. Thus, the ideal generated by $\mathfrak{a}$ and $\mathfrak{b}$ is simply the set of all sums $a + b$ and, consequently, is written as $\mathfrak{a} + \mathfrak{b}$. This general result could have been used in the proof that the residue classes modulo a maximal ideal form a field. The set $\{\alpha + a\beta\}$ in that proof can now be written as the ideal $\mathfrak{a} + a\mathfrak{o}$, which contains $a \notin \mathfrak{a}$ (use $\alpha = 0$ and $\beta = 1$) and, consequently, must be $\mathfrak{o}$.

2.8 Polynomial Rings. Let $\mathcal{R}$ be a ring with unit element. Then $a_o + a_1x + \dots + a_nx^n$ where $a_o, a_1, \dots, a_n \varepsilon \mathcal{R}$ is called a polynomial in x in the ring $\mathcal{R}$. With the customary rules of addition and multiplication as definitions of the operations, the set of all polynomials over $\mathcal{R}$ forms a ring $\mathcal{R}[x]$. A polynomial f should be viewed simply as a finite set of coefficients (elements) from $\mathcal{R}$ and written $f = (a_o, a_1, \dots, a_n)$. Then one can form the polynomial in x, $f(x) = a_o + a_1x + \dots + a_nx^n$ as well as say the polynomial in y, $f(y) = a_o + a_1y + \dots + a_ny^n$. Although a polynomial in x is not in general considered as a function of a variable ranging over $\mathcal{R}$, it may be so considered; for instance, for $a \varepsilon \mathcal{R}$ each $f(x) \varepsilon \mathcal{R}[x]$ can be mapped onto $f(a)$, giving a mapping of $\mathcal{R}[x]$ into $\mathcal{R}$ where $f(x) \overset{+}{\cdot} g(x) \longrightarrow f(a) \overset{+}{\cdot} g(a)$. In college algebra, polynomials are usually used as functions. For instance, the result from college algebra, that two polynomials taking on the same values over the reals are equal, is not true in general. Consider $\mathcal{R} = \{a_1, a_2, \dots, a_n\}$, let $f(x) = (x - a_1)(x - a_2) \dots (x - a_n) = x^n + \dots$, and let $g(x) = 0$ (the polynomial all of whose coefficients are zero); then $f(a_i) = g(a_i)$ for each $a_i \varepsilon \mathcal{R}$, but $f(x) \neq g(x)$. The knowledge of the division of polynomials in case $\mathcal{R}$ is a field is assumed to be familiar. The degree of a polynomial $a_o + a_1x + \dots + a_nx^n$ is defined as n if $a_n \neq 0$ and as $-\infty$ if $a_n = a_o = 0$. For $\mathcal{R}$ a domain of integrity, the degree of the product of two polynomials equals the sum of the degrees of the factors.

A polynomial in x and y is a sum of terms $a_{ij}x^iy^j$, $a_{ij} \varepsilon \mathcal{R}$. The polynomial ring in x and y, $\mathcal{R}[x, y]$ can be constructed by using the polynomials in $\mathcal{R}[x]$ as coefficients for polynomials in y, i.e. $\mathcal{R}[x, y] = \mathcal{R}[x][y]$. This definition is extended to polynomials with a finite number

of indeterminates in the obvious manner. If an infinite set of indeterminates is given, the totality of polynomials in all finite subsets of the infinite set forms a ring over $\mathfrak{o}$ in infinitely many indeterminates. Note, however, that each polynomial contains only a finite number of the indeterminates.

2.9 Fields. Let k be a field and f be a homomorphism of k into some ring, $f : k \longrightarrow$ some ring. The kernel $\mathfrak{n}$ of this homomorphism is an ideal of k. It may be that $\mathfrak{n} = \{0\}$. In this case, f is one-one as each coset contains a single element, i.e. f is an isomorphism into (homomorphism and one-one). Otherwise, there exists in $\mathfrak{n}$ an element $a \neq 0$ and, consequently, in k its inverse a^{-1}. However, $a \in \mathfrak{n} \Longrightarrow a^{-1}a = 1 \in \mathfrak{n} \Longrightarrow k \subset \mathfrak{n}$. Thus, $\mathfrak{n} = k$. It follows that the kernel $\mathfrak{n}$ of a homomorphism on a field k is either {0} or k. In the former case, the homomorphism is an isomorphism into, and, in the latter case, each element of the field maps onto zero. To exclude the latter case in applications, it is sufficient to show that at least one element of k, for instance the unit element, does not map onto zero. As any ring has the ring itself and {0} as ideals, these are called the trivial ideals of a ring. Hence, a field has only trivial ideals.

Let k be a (commutative) field. The next objective is to establish the existence of a field $K \supset k$ such that K is algebraically closed, i.e. such that every polynomial in K has a root in K. The seemingly weaker result, there exists a field $k_1 \supset k$ such that every polynomial with coefficients in k has a root in k_1, provides an infinite sequence of fields $k_1 \subset k_2 \subset \dots \subset k_n \subset \dots$ where every polynomial with coefficients in k_i has a root in k_{i+1}. $\bigcup_{n=1}^{\infty} k_n$ will be shown to fulfill the requirements of K

although by more advanced methods the k_1 introduced can be shown to be algebraically closed. Let $K = \bigcup_{n=1}^{\infty} k_n$ and let the coefficients of a given polynomial of K be $a_o, a_1, \ldots, a_n$, each of which is in a certain k_i; consequently, from the nested nature of the sequence there must be one of its members, say k_j, to which all of the coefficients $a_o, a_1, \ldots, a_n$ belong; thus the polynomial has a root in k_{j+1} and all the more in K. It will suffice to prove the existence of a field $\bar{k}_1$ containing an isomorphic replica $\bar{k}$ of k such that every polynomial in $\bar{k}$ has a root in $\bar{k}_1$, for k_1 can then be obtained from $\bar{k}_1$ by replacing $\bar{k}$ by k. Let f range over all polynomials $(a_o, a_1, \ldots, a_n)$ of k having $n \geq 1$ and $a_n = 1$. Let a variable x_f correspond to each of these polynomials f, and let $\mathfrak{P}$ be the polynomial ring in these x_f with coefficients in k (note that no f corresponds to any element of k but that $k \subset \mathfrak{P}$). Let $\mathfrak{A}$ be the ideal generated by <u>all</u> $f(x_f)$. Since $\mathfrak{A} \neq \mathfrak{P}$ (proof will be given later), the existence of a maximal ideal $\mathfrak{M}$ containing $\mathfrak{A}$ is assured. Consider now the canonical map, $p : \mathfrak{P} \longrightarrow \mathfrak{P}/\mathfrak{M}$, where $p(a) = a + \mathfrak{M}$. For brevity, let the field $\mathfrak{P}/\mathfrak{M} = \bar{k}_1$, $p(a) = \bar{a}$, and $p(k) = \bar{k}$. As $\mathfrak{M} \neq \mathfrak{P}$ ($\mathfrak{M}$ maximal), the homomorphism p restricted to k is an isomorphism. Finally, let $\bar{f} = (\bar{a}_o, \bar{a}_1, \ldots, \bar{a}_{n-1}1)$ be a polynomial in $\bar{k}$ where $f = (a_o, a_1, \ldots, a_{n-1}, 1)$ is the corresponding polynomial in k. $\bar{x}_f = p(x_f)$ is a root of $\bar{f}$, since $f(x_f) \in \mathfrak{A} \subset \mathfrak{M}$, kernel $\Longrightarrow p\left(f(x_f)\right) = 0 \Longrightarrow p(a_o + a_1 x_f + \ldots + a_{n-1} x_f^{n-1} + x_f^{n}) = 0 \Longrightarrow$ $\bar{a}_o + \bar{a}_1 \bar{x}_f + \ldots + \bar{a}_{n-1} \bar{x}_f^{n-1} + \bar{x}_f^{n} = 0$.

To complete the proof that K is an algebraic closure of k, it will now be proved indirectly that $\mathfrak{m} \neq \mathfrak{o}$. Otherwise, $1 = \Phi_1 f_1(x_1) + \dots + \Phi_n f_n(x_n)$ where $f_i(x_i)$ is an abbreviation for $f_i(x_{f_i})$ and Φ_i are polynomials in all x_f with coefficients in k. Assume n to be the smallest integer for which 1 can be so expressed. View the Φ_i as polynomials in x_1 with coefficients which are polynomials in the other x_f, and divide by $f_1(x_1)$ to get $\Phi_i = Q_i f_1(x_1) + R_i$ $(i = 2, 3, \dots, n)$ where the x_1-degree of R_i is less than the x_1-degree of $f_1(x_1)$. Consequently $1 = \Phi_1 f_1(x_1) + [Q_2 f_1(x_1) + R_2] f_2(x_2) + \dots$

$+ [Q_n f_1(x_1) + R_n] f_n(x_n) = [\Phi_1 + Q_2 f_2(x_2) + \dots + Q_n f_n(x_n)] f_1(x_1) +$

$[R_2 f_2(x_2) + \dots + R_n f_n(x_n)]$. On the one hand, $[\Phi_1 + Q_2 f_2(x_2) +$

$\dots + Q_n f_n(x_n)]$ must be zero to prevent the x_1-degree of

$[\Phi_1 + Q_2 f_2(x_2) + \dots + Q_n f_n(x_n)] f_1(x_1)$ from exceeding the x_1-degree of

$[R_2 f_2(x_2) + \dots + R_n f_n(x_n)]$; on the other hand,

$[\Phi_1 + Q_2 f_2(x_2) + \dots + Q_n f_n(x_n)] = 0$ contradicts the choice of n.

<u>Exercise A</u>: Let R be the additive group of real numbers; C, the additive group of complex numbers; R_+, the multiplicative group of positive real numbers; C^*, the multiplicative group of non-zero complex numbers; $U = \{e^{i\varphi}\}$, the multiplicative group of complex numbers on the unit circle. Study the following mappings by deciding which are homomorphisms, which are isomorphisms, by finding the kernels, etc.:

1) $R_+ \longrightarrow R$ given by the logarithm function;

2) $C^* \longrightarrow R_+$ by the absolute value;

3) $R \longrightarrow C^*$ by $e^{2\pi ix}$, $x \varepsilon R$;

4) $C^* \longrightarrow C^*$ by z^n;

5) $C \longrightarrow C^*$ by e^z.

<u>Exercise B</u>: Any commutative domain of integrity can be embedded in a field. Let $\mathscr{J}$ be a commutative domain of integrity (assumed associative but not necessarily with a unit element). Consider the set of formal pairs (fractions) a/b where a and b range over $\mathscr{J}$ and $b \neq 0$.

1) Show that the relation defined by $\frac{a}{b} \sim \frac{c}{d} \longleftrightarrow ad = bc$ is an equivalence relation.

2) With addition and multiplication defined respectively as $\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$ and $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$, show that the equivalence relation is stable under both addition and multiplication.

3) Prove that the equivalence classes form a ring and even a field F.

4) Show that this field contains an isomorphic replica of the ring $\mathscr{J}$ by considering the mapping $\mathscr{J} \longrightarrow F$ given by $a \longrightarrow$ equivalence class of ab/b.

<u>Exercise C</u>: In any textbook on modern algebra study the notion of a vector space over a field, linear dependence and independence, dimension of a space, and basis of a space. These concepts will be assumed in Chapter IV.

CHAPTER III

ELEMENTARY ARITHMETIC

3.1 <u>Principal ideal rings.</u> A ring $\mathfrak{o}$ is called a <u>principal ideal ring</u> provided $1 \varepsilon \mathfrak{o}$, $\mathfrak{o}$ is a domain of integrity, and every ideal $\mathfrak{m}$ in $\mathfrak{o}$ is principal, i.e. $\mathfrak{m} = d\mathfrak{o}$. Examples of principal ideal rings are

1) Z ;

2) $k[x]$, polynomials in one variable over a field;

3) $\{a + bi\}$, the Gaussian integers;

4) $\{a + b\omega\}$, where $\omega^2 + \omega + 1 = 0$;

5) $\{a + b\sqrt{-2}\}$;

6) $\{a + b \cdot \frac{1 + \sqrt{-7}}{2}\}$;

7) $\{a + b \cdot \frac{1 + \sqrt{-163}}{2}$;

8) $\{a + b\sqrt{2}\}$.

Note: In 3) – 8) a and b range over Z.

To show that every ideal $\mathfrak{m}$ in $\mathfrak{o} = k[x]$ is principal, consider a) $\mathfrak{m} = \{0\} = 0 \cdot \mathfrak{m}$; b) otherwise, for d(x) a polynomial of lowest degree in $\mathfrak{m}$ and $d(x) \neq 0$, $d(x) \cdot \mathfrak{o} \subset \mathfrak{m}$; and, for each $f(x) \varepsilon \mathfrak{m}$, $f(x) = q(x)\, d(x) + r(x)$, where the degree of r(x) is less than the degree of d(x); $r(x) = f(x) - q(x)\, d(x) \varepsilon \mathfrak{m}$; hence, as the degree of r(x) is less than the degree of d(x), $r(x) = 0$; therefore, $f(x) = q(x)\, d(x) \varepsilon d(x) \cdot \mathfrak{o}$; consequently, $\mathfrak{m} = d(x) \cdot \mathfrak{o}$. To show that the Gaussian integers form a principal ideal ring, a similar argument is used with the role of degree replaced by absolute value.

The main property of a principal ideal ring is the unique factorization into primes, which means that the results of elementary number

theory carry over. Moreover, in a principal ideal ring, every prime ideal is maximal and is generated by a prime element of the ring. For two non-zero ideals $a\mathcal{O}$ and $b\mathcal{O}$, $a\mathcal{O} \subset b\mathcal{O} \Longleftrightarrow a \in b\mathcal{O}$ $(1 \in \mathcal{O}) \Longleftrightarrow$ a is a multiple of $b \Longleftrightarrow b \mid a$ (b divides a). If $a\mathcal{O} = b\mathcal{O}$, then $b \mid a$ and $a \mid b$, and there exist two elements of $\mathcal{O}$, u_1 and u_2, such that $u_1 = a/b$ and $u_2 = b/a = u_1^{-1}$. A number u in the ring whose inverse is also in the ring is called a unit. The set of all units of $\mathcal{O}$ is a commutative group under multiplication for the product of two units and the inverse of a unit are also units. In Z, the units are ± 1 ; in $k[x]$, the non-zero constants; in the Gaussian integers, ± 1 and $\pm i$; in $\{a + b\omega\}$, ± 1, $\pm \omega$, $\pm \omega^2$; in examples 5), 6), and 7), ± 1 ; in 8) there are infinitely many units, viz. the integral powers of $1 + \sqrt{2}$. Elements that are indistinguishable in terms of divisibility are lumped together in the same ideal by the equivalence relation $a \sim b \Longleftrightarrow a/b$ a unit $\Longleftrightarrow a\mathcal{O} = b\mathcal{O}$. A G.C.D. of $a_1, a_2, \ldots, a_n \in \mathcal{O}$ is a common divisor of $a_1, a_2, \ldots, a_n$ that is divisible by any common divisor. In a principal ideal ring, d a g.c.d. of $a_1, a_2, \ldots, a_n \Longleftrightarrow d\mathcal{O}$ the smallest ideal containing every $a_i\mathcal{O} \Longleftrightarrow d\mathcal{O} = a_1\mathcal{O} + a_2\mathcal{O} + \ldots + a_n\mathcal{O}$. Note that for $\mathcal{O} = Z$ this gives $d = a_1x_1 + a_2x_2 + \ldots + a_nx_n$, $x_i \in Z$, a well known result of elementary number theory.

If u is a unit in the ring $\mathfrak{o}$, then each $a \varepsilon \mathfrak{o}$ has the trivial factorization $a = u(u^{-1}a)$. An element p of the ring is called a <u>prime</u> if it is a non-unit and if $p = ab$ implies that either a or b is a unit. The non-primes then consist of the units and all elements that can be written as the product of two factors neither of which is a unit.

Since every domain of integrity contains the two trivial prime ideals, {0} and the whole ring, the use of prime ideal in a principal ideal ring usually excludes these. Let $\mathfrak{p}$ be a (non-trivial) prime ideal in a principal ideal ring $\mathfrak{o}$. Hence, $\mathfrak{p} = p\mathfrak{o}$ and p is not a unit. If p had a non-trivial factorization $p = ab$, then $p|ab$; moreover, $p \nmid a$ and $p \nmid b$ (otherwise say $a = pc$, hence $p = pcb$, $1 = cb$, and b would be a unit); therefore, $ab \varepsilon \mathfrak{p}$, $a \not\varepsilon \mathfrak{p}$, and $b \not\varepsilon \mathfrak{p}$; consequently $p\mathfrak{o}$ would not be a prime ideal. This proves that every prime ideal $\mathfrak{p}$ in a principal ideal ring is generated by a prime p. Conversely, if p is a prime and $\mathfrak{p} = p\mathfrak{o}$, then $\mathfrak{p}$ is even a maximal ideal. Suppose $\mathfrak{p} \subset \mathfrak{m} = a\mathfrak{o}$; $p\mathfrak{o} \subset a\mathfrak{o} \Longrightarrow a|p \Longrightarrow p = ab$; since p is a prime, either a is a unit or b is a unit; in the former case $a\mathfrak{o} = \mathfrak{o}$, and in the latter case $p\mathfrak{o} = a\mathfrak{o}$. Consequently, every (non-trivial) prime ideal in a principal ideal ring is maximal and is generated by a prime p.

The general discussion of principal ideal rings will be interrupted for an example. Let $\mathfrak{o} = k[x, y]$, i.e. $\mathfrak{o}$ is a polynomial ring in two variables over a field k. Map each $f(x, y) \varepsilon \mathfrak{o}$ onto its constant term, i.e. $f(x, y) \dashrightarrow f(0, 0)$. Evidently, this mapping is a homomorphism for which the image of $\mathfrak{o}$ is the field k. The kernel $\mathfrak{p}$ is a maximal ideal, for the residue class ring, being isomorphic to k, is a field. Note that $\mathfrak{p}$ is the set of polynomials in $\mathfrak{o}$ without constant terms. If,

however, each $f(x, y)$ is mapped onto $f(0, y)$, the kernel $\mathfrak{q}$ is a prime ideal but not maximal, as the image $k[y]$ is a domain of integrity but not a field. Since every element of $\mathfrak{q}$ must be divisible by x, $\mathfrak{q}$ is the principal ideal $x\mathfrak{o}$. Observe that $\mathfrak{q} \subset \mathfrak{y}$. Hence, $k[x, y]$ is not a principal ideal ring, for one (non-trivial) prime ideal properly contains another. Since the polynomials in two variables have unique factorization and do not constitute a principal ideal ring, for unique factorization it is not in general necessary to have a principal ideal ring. In algebraic number theory, however, unique factorization occurs if and only if every ideal is principal.

<u>Exercise</u>: Show that $\mathfrak{y}$ as given above is not principal.

The question of unique factorization in a principal ideal ring will now be studied. Since it has been proved that a prime p generates a prime ideal, $p|ab$ and $p \nmid a \Longrightarrow p|b$, i.e. $ab \varepsilon p\mathfrak{o}$ and $a \not\varepsilon p\mathfrak{o} \Longrightarrow$ $b \varepsilon p\mathfrak{o}$. If $u_1p_1p_2\ldots p_r = u_2q_1q_2\ldots q_s$ where u_1, u_2 are units and p_i, q_j are primes, then $r = s$ and each p_i is the product of a unit and some q_j. p_1 divides $u_2q_1q_2\ldots q_s$, therefore p_1 divides some q_j, say q_1. Hence, $q_1\mathfrak{o} \subset p_1\mathfrak{o}$; but, as both are maximal, $q_1\mathfrak{o} = p_1\mathfrak{o}$ and $q_1 = up_1$, u a unit. Consequently, by substitution and cancellation a shorter relation is obtained, and the proof can be completed by induction. The possibility of an infinite number of prime factors of an element in a principal ideal ring will now be excluded. Let S be <u>any</u> non-empty set of ideals. This set is inductively ordered and has, therefore, maximal elements (not necessarily in the sense of maximal ideals) in the set. To prove that S is inductively ordered, let $\{\mathfrak{m}_\alpha\}$ be a totally ordered subset of S. Form the ideal $\mathfrak{m} = \bigcup_\alpha \mathfrak{m}_\alpha$. Since the ring is a principal ideal ring, it will

follow that $\mathfrak{a}$ is in S. $\mathfrak{a} = a\mathfrak{o}$, $a \in \mathfrak{a} = \bigcup_{\alpha} \mathfrak{a}_{\alpha} \Longrightarrow a \in \mathfrak{a}_{\beta}$ for some $\beta \Longrightarrow a\mathfrak{o} \subset \mathfrak{a}_{\beta}$; but, obviously, $\mathfrak{a}_{\beta} \subset a\mathfrak{o}$; so $\mathfrak{a} = \mathfrak{a}_{\beta}$. Another way of saying this is that in every increasing chain of ideals $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \mathfrak{a}_3 \subset \ldots$ in a principal ideal ring, a point is reached beyond which all ideals are equal; this is called the *chain condition*.

An element $a \neq 0$ in a principal ideal ring is either a unit, a prime, or a product of primes. Consider the set of all ideals $a\mathfrak{o}$ where this theorem is false for a. Let $a\mathfrak{o}$ be one of the maximal elements which this set contains. Since a is neither a unit nor a prime, a must have a non-trivial factorization, a = bc. However, $b|a \Longrightarrow a\mathfrak{o} \subset b\mathfrak{o}$, but $a\mathfrak{o} \neq b\mathfrak{o}$ since c is not a unit. Therefore, b is either a unit, a prime, or a product of primes. Similarly, c, and consequently a, is either a unit, a prime, or a product of primes. This is a contradiction; hence, the set considered must be empty.

3.2 *Unique factorization rings.* If unique factorization as described in the preceding section holds in a ring, it is called a *unique factorization ring*. Unique factorization rings of course include principal ideal rings, but are more general. An element a in a unique factorization ring can be written $a = up_1^{m_1} p_2^{m_2} \ldots p_r^{m_r}$ where u is a unit, each p_i is a prime, and $p_i \nmid p_j$ for $i \neq j$; this can easily be seen by introducing the necessary units in any factorization of a into primes. $d|a \Longrightarrow a = db$. Let d and b be factored into primes. As the product of these two factorizations must be a and as the factorization of a is unique,

$d = u_1 p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}$ $(n_i \leq m_i)$ gives the structure of a divisor of a.

An application of this to a common divisor d of $a_1 = u_1p_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$, $m_i \geq 0$, and $a_2 = u_2p_1^{k_1}p_2^{k_2}\dots p_r^{k_r}$, $k_i \geq 0$, gives $d = u_3p_1^{n_1}p_2^{n_2}\dots p_r^{n_r}$ where every $n_i \leq \min(m_i, k_i)$. d becomes the g.c.d. when each $n_i = \min(m_i, k_i)$.

Let us return now to polynomial rings and consider the extension of an onto homomorphism of any ring $\mathscr{r}$ to a homomorphism of $\mathscr{r}[x]$. Suppose $\phi : \mathscr{r} \longrightarrow \bar{\mathscr{r}}$, given by $\phi(a) = \bar{a}$, is an onto homomorphism with kernel $\mathscr{n}$. Hence, $\overline{a+b} = \bar{a} + \bar{b}$ and $\overline{ab} = \bar{a}\,\bar{b}$. Consider the mapping $\bar{\phi} : \mathscr{r}[x] \longrightarrow \bar{\mathscr{r}}[x]$ where $f(x) = a_0 + a_1x + \dots + a_nx^n \in \mathscr{r}[x]$ has as image $\overline{f(x)} = \bar{a}_0 + \bar{a}_1x + \dots + \bar{a}_nx^n \in \bar{\mathscr{r}}[x]$. Clearly, $\overline{f(x) + g(x)} = \overline{f(x)} + \overline{g(x)}$. Let $g(x) = b_0 + b_1x + \dots + b_mx^m = \sum b_jx^j$ and $f(x) \cdot g(x) = \sum c_kx^k$ where $c_k = \sum\limits_{i+j=k} a_ib_j$. Then $\overline{f(x)} \cdot \overline{g(x)} = \sum \bar{d}_kx^k$ where $\bar{d}_k = \sum\limits_{i+j=k} \bar{a}_i\bar{b}_j = \overline{\sum\limits_{i+j=k} a_ib_j} = \overline{c_k}$ and $\bar{\phi}$ is a homomorphism with kernel = $\{f(x) \mid \overline{f(x)} = 0\} = \{f(x) \mid \overline{a_i} = 0 \text{ for each } i\}$. This means a_i in $\mathscr{n}$ so that kernel = $\mathscr{n}[x]$. Two special cases of this extension will be of importance: 1) ϕ an isomorphism $\Longrightarrow \bar{\phi}$ also an isomorphism; 2) $\mathscr{n}$ is a prime ideal $\mathscr{p} \Longleftrightarrow \bar{\mathscr{r}}$ is a domain of integrity $\Longrightarrow \bar{\mathscr{r}}[x]$ also a domain of integrity; hence, $\mathscr{p}[x]$ is a prime ideal, and $f(x)\,g(x) \in \mathscr{p}[x]$, $f(x) \notin \mathscr{p}[x] \Longrightarrow g(x) \in \mathscr{p}[x]$. These statements are usually written more loosely as $f(x)\,g(x) \equiv 0 \pmod{\mathscr{p}}$, $g(x) \not\equiv 0 \pmod{\mathscr{p}} \Longrightarrow f(x) \equiv 0 \pmod{\mathscr{p}}$; moreover, in many cases, the mapping is not mentioned explicitly, but the polynomials in $\bar{\mathscr{r}}[x]$ are constructed from the polynomials in $\mathscr{r}[x]$ by taking their coefficients modulo $\mathscr{n}$. Assume now that $\mathscr{r}$ is a unique factorization ring

and p is a prime in $\mathfrak{o}$. Then $p\mathfrak{o}$ is a prime ideal since from $p|ab$, $p \nmid a \Longrightarrow p|b$ it follows that $ab \varepsilon p\mathfrak{o}$, $a \not\varepsilon p\mathfrak{o} \Longrightarrow b \varepsilon p\mathfrak{o}$. If $f(x) \varepsilon (p\mathfrak{o})[x]$, then each coefficient is divisible by p, and one agrees that $p|f(x)$. The following result of Gauss can now be stated:

$p \nmid f(x)$, $p \nmid g(x) \Longrightarrow p \nmid f(x)\, g(x)$.

A polynomial f(x) is called primitive if 1 is the g.c.d. of its coefficients, i.e. if no single prime divides all of its coefficients. By virtue of the above result of Gauss, the product of primitive polynomials is primitive.

The aim of what follows is to show that, if $\mathfrak{o}$ is a unique factorization ring, then $\mathfrak{o}[x]$ is also a unique factorization ring. Embed $\mathfrak{o}$ in its quotient field k. Then the polynomials in $\mathfrak{o}[x]$ and in $k[x]$ will be called, respectively, polynomials with integral coefficients and polynomials with rational coefficients. Let a be the g.c.d. of the coefficients of a polynomial $f(x) \varepsilon \mathfrak{o}[x]$; a is uniquely determined up to units as factors. Hence, $f(x) = a\, f_o(x)$ where $f_o(x)$ is primitive. Gauss called a the content of the polynomial. More generally, a polynomial $f(x) \varepsilon k\,[x]$ can be written as $a \cdot f_o(x)$ where $a \varepsilon k$ and $f_o(x)$ is primitive; this can be accomplished as in $\mathfrak{o}[x]$ after a common denominator is first factored out. The rational number a is called the content of f(x). As for uniqueness,

$a\, f_o(x) = b\, g_o(x)$, a and $b \varepsilon k$, $f_o(x)$ and $g_o(x)$ primitive $\Longrightarrow$ $(da)f_o(x) = (db)g_o(x)$ where d is a common denominator of a and b $\Longrightarrow$ $db = dau$ where u is a unit $\Longrightarrow b = au$. Consequently, the content of a polynomial is unique up to a unit factor. The content of the product of two polynomials is equal to the product of their contents, for

$f(x) = a\, f_o(x)$, $g(x) = b\, g_o(x)$, $f_o(x)$ and $g_o(x)$ primitive $\Longrightarrow$

$f(x)\ g(x) = ab\ f_o(x)\ g_o(x)$ where ab is the content of $f(x)\ g(x)$ as $f_o(x)\ g_o(x)$ is primitive.

The investigation of factoring a polynomial $f(x) \varepsilon k[x]$ can be greatly simplified. First, it may be assumed that $f(x) \varepsilon \mathcal{O}[x]$; otherwise, one can multiply by a common denominator of its coefficients since the elements of k are the units of $k[x]$. Secondly, the search for factors can be confined to $\mathcal{O}[x]$ since any factorization in $k[x]$ yields a factorization in $\mathcal{O}[x]$. For instance, suppose $f(x) = g(x)\ h(x)$ where $f(x) \varepsilon \mathcal{O}[x]$ and $g(x)$, $h(x) \varepsilon k[x]$. Further, let a and b be the contents of $g(x)$ and $h(x)$ respectively. Then $g(x) = a\ g_o(x)$ and $h(x) = b\ h_o(x)$ where $g_o(x)$ and $h_o(x)$ are primitive. Moreover, $f(x) = ab\ g_o(x)\ h_o(x)$ where ab is the content of $f(x)$ as $g_o(x)\ h_o(x)$ is primitive. Since $f(x)$, however, has integral coefficients, its content must be integral, i.e. $ab \varepsilon \mathcal{O}$. Therefore $f(x) = [ab\ g_o(x)][h_o(x)]$ is a factorization in $\mathcal{O}[x]$. To show that these remarks are not trivial, consider $\mathcal{O} = Z[\sqrt{-5}] = \{a + b\sqrt{-5}\}$ where $a, b \varepsilon Z$. Here, $2x^2 + 2x + 3$, which is equal to $\frac{1}{2}(2x + 1 + \sqrt{-5})(2x + 1 - \sqrt{-5})$, is reducible (factors) in $k[x]$, but irreducible in $\mathcal{O}[x]$. Of course, $Z[\sqrt{-5}]$ is not a unique factorization ring.

Eisenstein's theorem. Let $\mathcal{O}$ be a unique factorization ring and $f(x) = a_n x^n + a_{n-1}x^{n-1} + \dots + a_o \varepsilon \mathcal{O}[x]$. If there exists a prime $p \varepsilon \mathcal{O}$ such that

1) $p \nmid a_n$

2) $p | a_i$ for $i < n$

3) $p^2 \nmid a_o$

then $f(x)$ is irreducible in $k[x]$.

It suffices to prove that $f(x)$ is irreducible in $\mathcal{O}[x]$. Assume $f(x) = g(x)\,h(x)$ where $g(x), h(x) \in \mathcal{O}[x]$ and have positive degrees. Then $f(x) \equiv g(x)\,h(x) \pmod{p}$ and, consequently, $a_n x^n \equiv g(x)\,h(x) \pmod{p}$. Since the ring $\mathcal{O}/p\mathcal{O}[x]$ is a domain of integrity, the only possible factorizations of $a_n x^n$ have the form $bx^r \cdot cx^{n-r}$. Therefore, $g(x) \equiv bx^r \pmod{p}$ and $h(x) \equiv cx^{n-r} \pmod{p}$. As the degree of $g(x) \geq r$ and the degree of $h(x) \geq n - r$, the degree of $g(x)\,h(x) = n \geq r + (n - r) = n$; consequently, the inequality cannot hold in either case, and r = degree of $g(x) > 0$, $n - r$ = degree of $h(x) > 0$. It follows that $g(x) = bx^r$ + terms divisible by p and $h(x) = cx^{n-r}$ + terms divisible by p. This means the constant term of $g(x)\,h(x)$ is divisible by p^2, which is a contradiction.

Example A: For $\mathcal{O} = Z$, Eisenstein's theorem for $p = 2$ establishes the irreducibility of $x^n - 2$.

Example B: Let $f(x) = \dfrac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \ldots + x + 1$ where p is a prime number. Then $f(x + 1) = \dfrac{(x + 1)^p - 1}{x} = x^{p-1} + \binom{p}{1}x^{p-2} + \ldots + \binom{p}{p-1}$ is irreducible in $Z[x]$ by Eisenstein's theorem since

1) $p \nmid 1$;

2) $p \mid \binom{p}{i}$ for $1 \leq i \leq p - 1$ since $\binom{p}{i} = \dfrac{p(p-1)\ldots(p-i+1)}{i!}$

is an integer containing p as a factor;

3) $p^2 \nmid \binom{p}{p-1} = p$.

It follows that $f(x)$ is irreducible, for $f(x) = g(x)\,h(x) \Longrightarrow f(x + 1) = g(x + 1)\,h(x + 1)$.

Exercise: Prove $f(x)$ irreducible in $Z[x]$ for $f(x) = \frac{x^{p^r} - 1}{x^{p^{r-1}} - 1} =$

$x^{(p-1)p^{r-1}} + x^{(p-2)p^{r-1}} + \dots + x^{p^{r-1}} + 1.$

The final step toward the objective, $\mathcal{O}$ a unique factorization ring $\Longrightarrow$ $\mathcal{O}[x]$ a unique factorization ring, will be a study of the primes P of $\mathcal{O}[x]$. These are either constant primes, i.e. primes of $\mathcal{O}$, or irreducible polynomials in $k[x]$ (or $\mathcal{O}[x]$) which have to be primitive or otherwise $f(x) = a\, f_o(x)$ would be a factorization. To prove the uniqueness of factorization in $\mathcal{O}[x]$, one needs to first prove the lemma: P a prime, $P|AB$, $P \nmid A \Longrightarrow P|B$ where P, A, and B are elements of $\mathcal{O}[x]$. First, suppose $P = p$, a prime in $\mathcal{O}$. Hence, p | content of AB = content of A $\cdot$ content of B; but $p \nmid$ content of A. Therefore, p | content of B, and consequently $p|B$. Secondly, suppose P is a primitive polynomial which is irreducible, $P|AB$, and $P \nmid A$ where the division is understood to be in $\mathcal{O}[x]$. Obviously, $P|AB$ in $\mathcal{O}[x] \Longrightarrow P|AB$ in $k[x]$. Moreover, $P \nmid A$ in $\mathcal{O}[x] \Longrightarrow P \nmid A$ in $k[x]$. Otherwise, in $k[x]$ $A = PC$ where A, being an element of $\mathcal{O}[x]$, has an integer as its content, and the content of P is 1; therefore, the content of C is an integer and $C \varepsilon \mathcal{O}[x]$, which is a contradiction. Since factorization is unique in $k[x]$, $P|B$ in $k[x]$, and consequently in $\mathcal{O}[x]$, by the same reasoning as before. A polynomial $f(x)$ of $\mathcal{O}[x]$ can obviously be factored into primes as one can first factor out the content, factor this content, and then decompose the primitive part into irreducible primitive parts. The uniqueness of this factorization is then proved as usual.

More generally, $\mathcal{J}$ a unique factorization ring ⟹ $\mathcal{J}[x_1, x_2, \ldots, x_n]$ a unique factorization ring, for $\mathcal{J}$ a unique factorization ring ⟹ $\mathcal{J}[x_1]$ a unique factorization ring ⟹ $\mathcal{J}[x_1][x_2]$ a unique factorization ring, etc. For example, the set $Z[x_1, x_2, \ldots, x_n]$ of polynomials in n indeterminates over the integers is a unique factorization ring as is the set $k[x_1, x_2, \ldots, x_n]$ of polynomials over a field k.

In applications of Galois theory the factorization of a polynomial over a field is needed. For an illustration of some of the methods used, consider the factorization of $f(x) = x^5 - x + 1$ in $Q[x]$, which can be restricted, of course, to $Z[x]$. First, assume the existence of a linear factor of f(x). In this case, it can be assumed that $f(x) = (x - a)\ g(x)$, as the highest coefficient in each factor must be a unit. Since a must divide the constant term of f(x), $a = \pm 1$; but this is impossible, as neither +1 nor -1 is a zero of f(x). Consequently, no linear factor exists. Next, let $f(x) = g(x)\ h(x)$, say g(x) is quadratic, and let $g(x) = x^2 + ax + b$. We shall apply to f(x) the method of Kronecker for determining in a finite number of steps the factorization of a polynomial in $\mathcal{J}[x]$, where $\mathcal{J}$ has only a finite number of units and the elements of $\mathcal{J}$ itself are factorable in a finite number of steps. As $f(-2) = -29$, $f(-1) = 1$, $f(0) = 1$, $f(1) = 1$, and $f(2) = 31$, the corresponding values possible for g(x) are $g(-2) = \pm 1$ or ± 29, $g(-1) = \pm 1$, $g(0) = \pm 1$, $g(1) = \pm 1$, and $g(2) = \pm 1$ or ± 31. It is easily shown in this case that no such g(x) is possible; hence, f(x) is irreducible. In general, however, a finite number of g(x) may satisfy the requirements necessary at this

point; if so, each may be accepted or rejected by using long division. If none are accepted, then $f(x)$ is irreducible. Otherwise, two factors of $f(x)$ are obtained, and the method may be repeated on these factors until a factorization into primes is finally reached. Essential for the method to work is that a polynomial in one variable of degree n is uniquely determined by its values at n + 1 values for the variable. It is helpful to use values of x for which $f(x)$ does not contain many prime factors. Also, a study of the factorization modulo a prime is often advantageous. This method of Kronecker fails for the factorization of $f(x) = x^5 - 5x + 12$ in the polynomial ring over the extension field $Q(\alpha)$ of one of its roots α. Here, obviously, $f(x) = (x - \alpha)g(x)$, but the factorization of $g(x)$ is very difficult. It can be shown that $g(x)$ has two quadratic factors.

CHAPTER IV

FIELD EXTENSIONS

4.1 *Degree of extensions.* If E and F are fields such that $F \subset E$, F is said to be a *subfield* of E, and E is called an *extension field* of F. Henceforth, except for this paragraph, all fields will be assumed to be commutative and associative. E can be regarded as either a left vector space or a right vector space over the field F, for E is an additive group and the products of elements of F with elements of E satisfy all the laws required of the scalar multiplication. The dimension of E as a left vector space is called the left degree $[E:F]_L$ of E over F (or of F under E); similarly, the right degree $[E:F]_R$ of E over F is the dimension of E as a right vector space. Whether these degrees must be equal is an unsolved problem. for division rings

With E assumed to be commutative, the distinction between right and left is not necessary, and the *degree of the extension* E over F is written $[E:F]$. Hence, $[E:F]$ is the maximum number of elements of E that are linearly independent over F, if a maximum number exists; otherwise, $[E:F]$ is said to be infinity. $\alpha_1, \alpha_2, \ldots, \alpha_n \varepsilon E$ are *linearly independent* if and only if $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$, $c_i \varepsilon F \Longrightarrow$ all $c_i = 0$. $\alpha_1, \alpha_2, \ldots, \alpha_n \varepsilon E$ are *linearly dependent* if and only if there exist $c_1, c_2, \ldots, c_n \varepsilon F$, not all zero, such that $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$. Observe that a single element α_1 of E is linearly independent if and only if $\alpha_1 \neq 0$, and consequently $[E:F] \geq 1$ since every field contains a non-zero element.

Suppose $[E:F] = 1$. $1 \varepsilon E$ is linearly independent. Therefore, 1 and any other element α of E are linearly dependent, i.e. $c\alpha + d = 0$ where $c, d \varepsilon F$ and $c \neq 0$. Hence, $\alpha = -c^{-1}d \varepsilon F$; so $E = F$. In general, if the set $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a maximum set of linearly independent elements in E, it is called a basis; further, $E = F\alpha_1 + F\alpha_2 + \ldots + F\alpha_n$. Again for $[E:F] = 1$, $1 \varepsilon E$ is linearly independent and constitutes a basis ; hence $E = F \cdot 1 = F$.

If F, E, and K are fields such that $F \subset E \subset K$, then $[K:F] = [E:F][K:E]$. Three cases will be distinguished.

1) Assume $[K:E]$ infinite. Hence, there exist arbitrarily many elements of K that are linearly independent with respect to E. But, elements that are linearly independent with respect to E are obviously linearly independent with respect to the subfield F. Therefore, $[K:F]$ is also infinite.

2) $[E:F] = \infty$. Since each set of elements of E that are linearly independent with respect to F is also a set of elements of K that are linearly independent with respect to F, $[K:F]$ is infinite.

3) $[K:E]$ and $[E:F]$ are both finite. Let $[K:E] = m$ and $[E:F] = n$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be a basis of E/F (E as an extension of F), i.e. $\alpha_1, \alpha_2, \ldots, \alpha_n$ are linearly independent with respect to F and any $\theta \varepsilon E$ can be uniquely expressed in terms of them, $\theta = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n$, $c_i \varepsilon F$. Similarly, for any $A \varepsilon K$, $A = \theta_1A_1 + \theta_2A_2 + \ldots + \theta_mA_m$ where $\theta_i \varepsilon E$ and $A_1, A_2, \ldots, A_m$ is a basis of K/E. As each $\theta_i \varepsilon E$, $\theta_i = c_{i1}\alpha_1 + c_{i2}\alpha_2 + \ldots + c_{in}\alpha_n$, $c_{ij} \varepsilon F$. Consequently, by substitution, $A = \sum\limits_{i,j} c_{ij}\alpha_jA_i$. Since each α_jA_i is in K, an arbitrary

element of K can be expressed as a linear combination of the mn elements $\alpha_j A_i$ with coefficients in F, and $[K:F] \leq mn$. To establish the equality, we prove these mn elements linearly independent with respect to F:

$$\sum_{i,j} c_{ij}\alpha_j A_i = 0,\ c_{ij} \in F \Longrightarrow \sum_{i=1}^{m}\left(\sum_{j=1}^{n} c_{ij}\alpha_j\right)A_i = 0 \Longrightarrow \sum_{j=1}^{n} c_{ij}\alpha_j = 0$$

$$\left(\sum_{j=1}^{n} c_{ij}\alpha_j \in E \text{ and } A_1, A_2, \ldots, A_m \text{ linearly independent with respect to } E\right)$$

$\Longrightarrow$ each $c_{ij} = 0$ ($c_{ij} \in F$ and $\alpha_1, \alpha_2, \ldots, \alpha_n$ linearly independent with respect to F). It should be remarked that an extension field E of a field F is called a finite extension when $[E:F]$ is finite.

4.2 Adjunction; simple field extensions. Let E be an extension field of F, S be a subset of E, and F(S) be the intersection of all fields (e.g. E) containing F and S. F(S) is clearly the smallest extension field of F that contains S, for the intersection is a field containing both F and S and must be contained in any field containing F and S. F(S) is said to be obtained by adjunction of S to F. All elements of E that can be expressed in terms of the elements of $F \cup S$ with the use of a finite number of rational operations must belong to F(S); but these elements form a field; hence, the totality of these elements is exactly F(S).

In case S contains only the element α, F(S) is written as $F(\alpha)$ and is called a simple field extension of F. $F(\alpha)$ consists of all $\frac{f(\alpha)}{g(\alpha)}$ where $f(x), g(x) \in F[x]$ and $g(x) \neq 0$, for $\frac{f(\alpha)}{g(\alpha)} \in F(\alpha)$ and the totality of all rational functions of α, $\frac{f(\alpha)}{g(\alpha)}$, is a field. Consider

the mapping $\varphi : F[x] \xrightarrow{\text{into}} F(\alpha)$ where $f(x) \longrightarrow f(\alpha)$. φ is a homomorphism having as kernel $\mathcal{J}$ the set of polynomials $f(x)$ for which $f(\alpha) = 0$. As the image $F[\alpha]$ is a subset of a field, it must be a domain of integrity; hence, $\mathcal{J}$ is a prime ideal of the principal ideal ring $F[x]$. $\mathcal{J} \neq F[x]$ for 1 does not map onto 0. The two remaining possibilities follow:

1) $\mathcal{J} = \{0\}$. As the map, in this case, is one-one into, $F[x]$ is mapped isomorphically into $F(\alpha)$ with image $F[\alpha]$, i.e. $F[x] \simeq F[\alpha]$. As the domains of integrity, $F[x]$ and $F[\alpha]$, are isomorphic, it must follow that their quotient fields, $F(x)$ and $F(\alpha)$ respectively, are also isomorphic, i.e. $F(\alpha)$ is isomorphic to the field of rational functions of a single indeterminate. α, in this case, is called transcendental with respect to the field F. It may be remarked that algebra does not distinguish between the extensions of two transcendental elements, e.g. $Q(e) \simeq Q(\pi) \simeq$ ~~$Q[x]$~~, $Q(x)$.

2) $\mathcal{J} \neq \{0\}$. As $F[x]$ is a principal ideal ring, $\mathcal{J}$ is generated by a prime element (irreducible polynomial) of $F[x]$, which is determined up to a unit (an element of F). Let the irreducible polynomial that generates $\mathcal{J}$ and has highest coefficient equal to 1 be called $p(x)$. Among all the polynomials in $F[x]$ having highest coefficient 1 and α as a zero, $p(x)$ is characterized by being irreducible <u>or</u> by having lowest degree. $p(x)$ is denoted by Irr (α, F). Consider now the canonical decomposition of the homomorphism   :

$$F[x] \longrightarrow F[x]/\mathcal{J} \xrightarrow{\simeq} \text{image}, F[\alpha] \xrightarrow{i} F(\alpha) .$$

As $\mathcal{J}$ is a prime ideal of a principal ideal ring, $\mathcal{J}$ is even maximal, and consequently $F[x]/\mathcal{J}$ is a field. Hence, its isomorphic image $F[\alpha]$ is

also a field. Moreover, i must be the identity mapping, for $F(\alpha)$ was the smallest field containing F and α. Consequently, $F[\alpha]$ is already the field obtained by adjunction of α to F and is isomorphic to the residue class ring (field) of $F[x]$ modulo the ideal generated by $p(x)$. In the isomorphism $F[x] / \mathcal{Y} \simeq F(\alpha)$, a residue class of $F[x] / \mathcal{Y}$, $f(x) + \mathcal{Y}$, is mapped onto $f(\alpha)$, and this mapping is one-one. A one-one mapping of a subset of $F[x]$ onto $F(\alpha)$ can be obtained from this isomorphism by selecting a unique representative $r(x)$ from each residue class, $f(x) + \mathcal{Y}$. $r(x)$ is obtained from $f(x)$ by the long division $f(x) = p(x)\, q(x) + r(x)$ where the degree of $r(x)$ is less than the degree of $p(x)$. Hence, $f(x) \equiv r(x) \pmod{p(x)}$, and $f(x) + \mathcal{Y} = r(x) + \mathcal{Y}$. As for uniqueness, $r_1(x) \equiv r_2(x) \pmod{p(x)} \Longrightarrow r_1(x) - r_2(x) \equiv 0 \pmod{p(x)} \Longrightarrow r_1(x) - r_2(x) = 0$, for the degree of $r_1(x) - r_2(x)$ is less than the degree of $p(x)$. The mapping given by $r(x) \longrightarrow r(\alpha)$ is one-one and onto, but not a homomorphism, as $r_1(x) \cdot r_2(x)$ will not in general be the representative selected from the residue class $r_1(x)\, r_2(x) + \mathcal{Y}$. Consequently, if the degree of $p(x)$ is n, then every $\theta \in F(\alpha)$ is of the form $\theta = r(\alpha)$, i.e. $\theta = a_0 + a_1\alpha + \dots + a_{n-1}\alpha^{n-1}$ with $a_i \in F$. This expression for θ is unique; otherwise, α would satisfy an equation with degree lower than the degree of $p(x)$. Therefore, $1, \alpha, \alpha^2, \dots, \alpha^{n-1}$ is a basis for the extension $F(\alpha)$ over F, and $[F(\alpha):F] = n$. Incidentally, in the classical approach, the degree of $F(\alpha)$ over F was defined as the degree of the irreducible polynomial $p(x)$ for which $p(\alpha) = 0$. This definition necessitates a demonstration that the degree is a property of the field, i.e. if the same field is generated by two different elements α and β, then the degree is independent of the choice of α or β. Another bad feature of the old definition is the difficulty encountered in showing the

multiplicative behavior of the degrees for successive simple extensions, $F \subset F(\alpha) \subset F(\alpha, \beta)$; as a matter of fact the part of the classical proof showing that $F(\alpha, \beta)$ is a simple extension of F is not even true for an abstract field F.

Example: Let F be Q, (rationals) p(x) be the polynomial $x^5 - x + 1$ (irreducible over Q), and α be an element from an algebraic closure of Q such that $p(\alpha) = 0$. If $\theta \in Q(\alpha)$, then $\theta = a_0 + a_1\alpha + \dots + a_4\alpha^4$ (unique). For instance, θ might be $2 - 3\alpha + 4\alpha^3 - 2\alpha^4$. Since uniqueness means that no other polynomial of the fourth degree in α can equal θ, the elements of $Q(\alpha)$ can be enumerated. Addition and subtraction are carried out in the usual way, and multiplication is performed as for polynomials, but then reducing by dividing by $p(\alpha)$. To carry out several multiplications it is useful to make the table:

$$\alpha^5 = \alpha - 1$$
$$\alpha^6 = \alpha^2 - \alpha$$
$$\alpha^7 = \alpha^3 - \alpha^2$$
$$\alpha^8 = \alpha^4 - \alpha^3$$

Observe that, in general, the degree of each member on the right can be kept less than n by using the first equation when terms of higher degree occur. For division, the method of undetermined coefficients may be used, e.g. $\frac{1}{2 - 3\alpha + 4\alpha^3 - 2\alpha^4} = x_0 + x_1\alpha + \dots + x_4\alpha^4 \Longrightarrow 1 = (2 - 3\alpha + 4\alpha^3 - 2\alpha^4)(x_0 + x_1\alpha + \dots + x_4\alpha^4) \Longrightarrow$ (by multiplying and using the table) $1 = \ell_0 + \ell_1\alpha + \dots + \ell_4\alpha^4$ where each ℓ_i is linear in $x_0, x_1, \dots, x_4$. Since 1 is uniquely represented, $x_0, x_1, \dots, x_4$ can be determined by solving the system of linear equations: $\ell_0 = 1$, $\ell_1 = 0$,

$\ell_2 = 0, \ldots, \ell_4 = 0$. These equations in 5 unknowns have a unique solution since it is known that $Q(\alpha)$ is a field. It should be noted that this division is exactly what was called "rationalizing the denominator" in elementary algebra.

Exercise A: As in the example above, take $p(x) = x^3 - x + 1$ and compute $\dfrac{1}{1 - 2\alpha + 3\alpha^2}$.

Exercise B: Introduce the complex number field by adjunction of i to the real number field R by taking $p(x) = x^2 + 1$ and $p(i) = 0$.

If E/F is a finite extension, i.e. if $[E:F] = n < \infty$, then any $\theta \in E$ is algebraic over F. (To say an element is algebraic over F means that it is a zero of a polynomial over F.) Proof: $1, \theta, \theta^2, \ldots, \theta^n$ are $n + 1$ elements of E and hence are linearly dependent; so, $c_0 + c_1\theta + \ldots + c_n\theta^n = 0$, $c_i \in F$, not all $c_i = 0$.

A method known as the Tschirnhausen transformation will be indicated for finding a polynomial having as a zero the element $\theta = 2 - 3\alpha + 4\alpha^3 - 2\alpha^4$ of the quintic field in the example above; $p(x) = x^5 - x + 1$ and $F = Q$. $\theta^0, \theta^1, \theta^2, \theta^3, \theta^4, \theta^5$ can be computed in terms of $\alpha, \alpha^2, \alpha^3, \alpha^4$; elimination of α from the resulting equations yields a polynomial having θ as a zero. It would probably be simpler to compute $\theta, \alpha\theta, \alpha^2\theta, \alpha^3\theta, \alpha^4\theta$ in terms of $\alpha, \alpha^2, \alpha^3, \alpha^4$. Since the resulting system of five homogeneous linear equations in $1, \alpha, \alpha^2, \alpha^3, \alpha^4$ has a non-trivial solution, its determinant must be zero. This yields a polynomial equation of the fifth degree that has θ as a root. The fifth degree polynomial so obtained is irreducible in this case for by the multiplicative property of the degrees the degree of $Q(\theta)$

is either 5 or 1 and 1 is impossible as θ is not rational. In most cases, only the constant term, called the norm, of this polynomial is needed.

Exercise C: Find an irreducible polynomial having $\theta = 1 + \alpha - 2\alpha^2$ as a zero, if $p(x) = x^3 - x + 1$, $F = Q$, and $p(\alpha) = 0$.

Let E be an extension field of F, $F \subset E$; let E_o be the extension field obtained from F by adjunction of $\alpha_1, \alpha_2, \ldots, \alpha_r \varepsilon E$, i.e. $E_o = F(\alpha_1, \alpha_2, \ldots, \alpha_r)$. Contention: If $\alpha_1, \alpha_2, \ldots, \alpha_r$ are algebraic over F, then E_o is a finite extension of F, and every element in E_o is a polynomial in $\alpha_1, \alpha_2, \ldots, \alpha_r$ with coefficients in F. Before the proof is given, notice that operations on α_i and α_j are defined only if they belong to a comprehending set in which these operations are already defined; furthermore, observe from the conclusion that quotients of polynomials will not be needed in E_o.

Proof by induction on r:

1) For r = 0, the contention trivially follows.

2) Put $E' = F(\alpha_1, \alpha_2, \ldots, \alpha_{r-1})$. Then $E_o = E'(\alpha_r)$. $[E_o:E']$ is finite since α_r, being algebraic over F, is algebraic over E'. $[E':F]$ is finite by the inductive assumption. Consequently, $[E_o:F] = [E_o:E'][E':F]$ is finite. Let $\theta \varepsilon E_o$, then $\theta = \phi(\alpha_r)$ where ϕ is a polynomial with coefficients in E'; but each coefficient is a polynomial in $\alpha_1, \alpha_2, \ldots, \alpha_{r-1}$ with coefficients in F; therefore, θ is a polynomial in $\alpha_1, \alpha_2, \ldots, \alpha_r$ with coefficients in F. Conversely, if E_o is a finite extension of F, then E_o can be obtained by adjunction of a finite number of elements of F. For instance, the adjunction of all elements in a basis of E_o over F suffices.

An extension E/F is called *algebraic* if *every* element of E is algebraic over F; otherwise, an extension is called *transcendental*. It has been shown that every finite extension field is algebraic. It will now be shown that a tower of algebraic extensions is algebraic. Let E_2 be an algebraic extension of E_1 and E_1 be an algebraic extension of F; then E_2 is an algebraic extension of F. Proof: Let $\theta \epsilon E_2$, then there exists an equation $\theta^n + \alpha_1\theta^{n-1} + \dots + \alpha_n = 0$ where all $\alpha_i \epsilon E_1$. Consider $E_3 = F(\alpha_1, \alpha_2, \dots, \alpha_n)$ and $E_4 = E_3(\theta)$. E_3/F is finite by the theorem of the preceding paragraph, and E_4/E_3 is finite since θ satisfies an equation with coefficients in E_3. Hence, E_4/F is a finite extension, and consequently θ is algebraic over F.

4.3 *Ruler and compass constructions.* The permissible basic ruler and compass operations are a free choice of a point within a certain region and constructions of a straight line through two points, a circle with center and radius given, the intersection of two straight lines, the intersection of a straight line and a circle, and the intersection of two circles. From a given set of points, lines, and circles, the general problem is to construct another specified set by means of a finite number of basic operations. Consider, for instance, the problem of duplicating a cube, i.e. given an edge of a cube, to construct the edge of another cube whose volume is twice that of the first one. This problem is equivalent to locating the point $(\sqrt[3]{2}, 0)$ in the coordinate plane given the unit interval from $(0, 0)$ to $(0, 1)$. Assume such a construction is possible. At each stage of the construction where a free choice of a point is available, the coordinates chosen can be assumed to be rational.

Moreover, each basic step in the construction yields geometric elements having coordinates that can be expressed in terms of rational operations and square roots of the coordinates existing up to that step. Those new coordinates then are in an extension field of degree 2 or 1 over the field containing the old coordinates. As the construction must end in a finite number of steps, the extension field E finally reached must have degree 2^m over the ground field of rational numbers, Q. If the construction is to be successful, $\sqrt[3]{2}$ must be in this field, and $Q(\sqrt[3]{2})$ must be a field intermediate to Q and E. Since $\sqrt[3]{2}$ is a zero of $x^3 - 2$ irreducible over Q, the degree of $Q(\sqrt[3]{2})$ over Q must be 3. Since this contradicts $[E:Q] = [E:Q(\sqrt[3]{2})][Q(\sqrt[3]{2}):Q]$, the construction is impossible. In general, each coordinate of a geometrical element that can be constructed with ruler and compass must be a zero of a polynomial over Q with a power of 2 as its degree.

Exercise: Use $\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta$ to show that an angle of 60° cannot be trisected.

4.4 Intermediate Fields. Suppose $E = F(\alpha)$ where α is algebraic over F. Contention: There exist only a finite number of fields E_0 intermediate between F and E, $F \subset E_0 \subset E$. Proof: Let P(x) be the irreducible polynomial over F having α as a zero, i.e. $P(x) = \text{Irr}(\alpha, F)$. (We shall restrict ourselves to polynomials with highest coefficient 1.) Associate with E_0 the polynomial $K(x) = \text{Irr}(\alpha, E_0)$, and let the degree of K(x) be r. As $E = E_0(\alpha)$, $[E:E_0] = r$. Adjoin the coefficients of K(x) to F to obtain a field E_1: $F \subset E_1 \subset E_0$. On the one hand, from the manner in which E_1 was constructed, $K(\alpha) = 0$ in E_1, and $\text{Irr}(\alpha, E_1)$ has degree at most r;

hence, $[E:E_1] \leq r$ since $E = E_1(\alpha)$. On the other hand, $[E:E_1] \geq r$ since $E_1 \subset E_0 \subset E$. Therefore, $[E:E_1] = r$. Consequently, $[E_0:E_1] = 1$, and $E_0 = E_1$. This means that the correspondence between all such E_0 and all such $K(x)$ is one-one, and the proof may be completed by showing that only a finite number of possibilities exist for $K(x)$. As $P(\alpha) = 0$ and $K(x)$ is the polynomial over E_0 of lowest degree that has α as a zero, $K(x) | P(x)$ in E_0 and certainly in E. But, in E, $P(x)$ has as factors only a finite number of irreducible polynomials (e.g. $x - \alpha$) with leading coefficient 1. It follows, therefore, that the number of possibilities for $K(x)$ is finite.

Example: Let $E = Q(\sqrt{2}, \sqrt{3})$. Since $\sqrt{2}$ and $\sqrt{3}$ satisfy quadratic equations, each of $[Q(\sqrt{2}):Q]$ and $[E:Q(\sqrt{2})]$ is either 1 or 2; so $[E:Q] = 1, 2,$ or 4. As $\sqrt{2}$ can be proved irrational, only 2 or 4 is possible. If $[E:Q]$ were 2, then $[E:Q(\sqrt{2})] = 1$ and $\sqrt{3} = a + b\sqrt{2}$, a and b ε Q $\Longrightarrow 3 = a^2 + 2b^2 + 2ab\sqrt{2} \Longrightarrow$ (in $Q(\sqrt{2})$ an element is uniquely expressed as a linear polynomial in $\sqrt{2}$) $a^2 + 2b^2 = 3$ and $2ab = 0 \Longrightarrow a = 0$ or $b = 0$. For $a = 0$, $\sqrt{3} = b\sqrt{2}$, and $\sqrt{6} = 2b$; for $b = 0$, $\sqrt{3} = a$. A contradiction is obtained in each case since $\sqrt{6}$ and $\sqrt{3}$ can be proved irrational. Hence, if $\theta \varepsilon E$, $\theta = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$. Let $\alpha = \sqrt{2} + \sqrt{3}$. Then $\alpha^2 = 5 + 2\sqrt{6}$ and $\alpha^3 = 11\sqrt{2} + 9\sqrt{3}$. Since it follows that

$$\sqrt{2} = \frac{\alpha^3 - 9}{2} \quad \text{and} \quad \sqrt{3} = \frac{11\alpha - \alpha^3}{2}, \quad E = Q(\alpha).$$

Thus, any element of E can also be written as $a + b\alpha + c\alpha^2 + d\alpha^3$. Furthermore, by squaring $\alpha^2 - 5 = 2\sqrt{6}$, $\alpha^4 - 10\alpha^2 + 1 = 0$ is obtained. Hence, $\text{Irr}(\alpha, Q) = x^4 - 10x^2 + 1$. To illustrate how to find all fields

E_0 intermediate to E and Q by the method of the preceding paragraph, let $P(x) = x^4 - 10x^2 + 1$. Observe that for $\beta = \sqrt{2} - \sqrt{3}$, $\gamma = -\sqrt{2} + \sqrt{3}$, and $\delta = -\sqrt{2} - \sqrt{3}$, β, γ, and δ are also zeros of P(x). Therefore, $P(x) = (x - \alpha)(x - \beta)(x - \gamma)(x - \delta)$ in E. For E_0 different from Q or E, the degree of the polynomial K(x) must be 2. Hence, as $K(\alpha) = 0$, the possibilities for K(x) are:

$$(x - \alpha)(x - \beta) = x^2 - 2\sqrt{2}\, x - 1 ,$$
$$(x - \alpha)(x - \gamma) = x^2 - 2\sqrt{3}\, x + 1 ,$$
$$(x - \alpha)(x - \delta) = x^2 - (5 + 2\sqrt{6}) .$$

The corresponding possibilities for E_0 ($= E_1$) are $Q(\sqrt{2})$, $Q(\sqrt{3})$, and $Q(\sqrt{6})$, obtained respectively by adjoining to Q the coefficients of the K(x) listed. Hence, there are 5 fields intermediate to E and Q in this case, viz. E, $Q(\sqrt{2})$, $Q(\sqrt{3})$, $Q(\sqrt{6})$, and Q. For ordinary number fields, as in this example, a finite extension field can be generated by one element; however, this is not true in general.

The converse of the theorem that there exist only a finite number of intermediate fields to F and $F(\alpha)$ where α is algebraic over F will be given in this paragraph in case F has infinitely many elements. The discussion for the finite case will be given later. Let F have infinitely many elements, and let E be a finite extension of F such that there are only a finite number of fields intermediate to E and F. Contention: E can be generated by one element. Proof: Assume first that E is generated from F by two elements, $E = F(\alpha, \beta)$. Let c range over the infinitely many elements of F, let $\gamma_c = \alpha + c\beta$ correspond to each c, and consider the fields $F(\gamma_c)$. Since these fields are intermediate to E and F, only a finite number of them can be different, by the hypothesis. This means that there

exist $c, d \varepsilon F$ such that $c \neq d$ and $F(\gamma_c) = F(\gamma_d)$. In particular, $\alpha + c\beta$, $\alpha + d\beta \varepsilon F(\gamma_c) \longrightarrow (d - c)\beta \varepsilon F(\gamma_c) \longrightarrow$ (since $d - c$ is a non-zero element of F) $\beta \varepsilon F(\gamma_c) \longrightarrow c\beta \varepsilon F(\gamma_c) \longrightarrow \alpha \varepsilon F(\gamma_c)$. However, if $\alpha, \beta \varepsilon F(\gamma_c)$, then $E = F(\alpha, \beta) \subset F(\gamma_c)$; but, obviously, $F(\gamma_c) \subset E$. Therefore, $E = F(\gamma_c)$. The proof is completed by induction. Let $E = F(\alpha_1, \alpha_2, \ldots, \alpha_r)$ and $E' = F(\alpha_1, \alpha_2, \ldots, \alpha_{r-1})$. The inductive assumption gives $E' = F(\gamma)$ for some $\gamma \varepsilon E'$. Hence, $E = F(\gamma, \alpha_r) = F(\delta)$ for some $\delta \varepsilon E$ by the case $r = 2$.

<u>Exercise</u>: Consider the residue class field modulo 2, $k = \{0, 1\}$. Let $E = k(x, y)$ be the rational functions of x and y with coefficients in k. Let $F = k(x^2, y^2)$ be the rational functions of x^2 and y^2 with coefficients in k. Prove that the degree of E over F is 4 and that the square of every element of E is an element of F. Hence, E cannot be obtained from F by adjoining one element, and consequently there must be an infinite number of intermediate fields.

Theorem: If F is a field and G is a finite multiplicative subgroup of F, $G \neq \{0\}$, then G is cyclic, i.e. G consists of the powers of a single element. That F is a field is not used in the first part of the proof.

1) Let G be any finite commutative group. Let $A, B \varepsilon G$ and a, b be their respective periods (orders), i.e. $A^a = 1$, $B^b = 1$ where a and b are the smallest positive integers for which this is true. Let m be the least common multiple of a and b. m can be written as the product of two relatively prime factors, $m = (p_1^{u_1} p_2^{u_2} \ldots p_r^{u_r})(q_1^{v_1} q_2^{v_2} \ldots q_s^{v_s})$ where $p_i^{u_i} \mid a$ and $q_j^{v_j} \mid b$. Put $c = p_1^{u_1} p_2^{u_2} \ldots p_r^{u_r}$,

$d = q_1^{v_1} q_2^{v_2} \ldots q_s^{v_s}$, and $C = A^d B^c$. C is an element of period m in G: for $C^m = (A^m)^d (B^m)^c = 1$; and $C^x = 1 \Longrightarrow A^{xd} B^{xc} = 1 \Longrightarrow (A^{xd} B^{xc})^d = A^{xd^2} = 1 \Longrightarrow$ (A has period a) $a|xd^2 \Longrightarrow$ (since $c|a$) $c|xd^2 \Longrightarrow$ (since c and d are relatively prime) $c|x$, and similarly $d|x \Longrightarrow m|x$. Hence, one can construct an element of G whose period is the least common multiple of the periods of two given elements of G. Contention: If, in particular, B is an element of G whose period b is largest, then $A^b = 1$ for <u>all</u> $A \varepsilon G$. The proof is immediate, for C as constructed above has as period m, the least common multiple of a and b; but by the choice of B, $m = b$; therefore $A^b = A^m = 1$ since $a|m$. A counterexample in the case of non-abelian groups is the symmetric group of all permutations on 3 elements; this group of order 6 has elements only of orders 1, 2, and 3.

2) If F is a field, $f(x) \varepsilon F[x]$, and the degree of $f(x) = n \geq 0$, then $f(x) = 0$ has at most n distinct roots in F. The result is trivial for $n = 0$. Hence, let $f(x) = k P_1(x) P_2(x) \ldots P_r(x)$ where k is the content, each $P_i(x)$ is irreducible over F, and $r \leq n$. If the degree of $P_i(x)$ exceeds 1, it cannot happen that $P_i(c) = 0$ for $c \varepsilon F$; otherwise $x - c \mid P_i(x)$, which is a contradiction. Therefore, the roots of $f(x) = 0$ are the roots of those $P_i(x) = 0$ where $P_i(x)$ has degree 1. Clearly then, the number of roots of $f(x) = 0$ is at most r, and consequently at most n. Note, however, that there exists a group of order 4 such that $x^2 = 1$ for every x.

3) Now let $G \neq \{0\}$ be a finite multiplicative group of order n in a field F. Let the period b of B be the largest period occurring in G. Then, by 1), $x^b = 1$ for <u>all</u> $x \varepsilon G$, i.e. the equation $x^b = 1$ has at least n roots. Hence, $n \leq b$ by 2). On the other hand, as the powersof $x \varepsilon G$

form a subgroup of G, $x^n = 1$ for all $x \varepsilon G$, and $n \geq b$ since there is an element of G with period b. This shows $b = n$. So there exists an element B whose period is n. This means that 1, B, B^2, ..., B^{n-1} are distinct elements; moreover, as there are n of them, they are the elements of G. These elements, being precisely the solutions of the equation $x^n = 1$, are the n^{th} roots of unity.

Corollary: Suppose F is a finite field with q elements, then the elements $\neq 0$ form a cyclic group G under multiplication of order $q - 1$, i.e. there exists an element $g \varepsilon F$ such that $G = F - \{0\} = \{1, g, g^2, \ldots, g^{q-2}\}$ and $g^{q-1} = 1$. In number theory Z/pZ is a field with p elements; hence, the non-zero elements form a cyclic group of order $p - 1$, e.g. for $p = 17$ a primitive element can be found by experiment (3 is one).

The converse of the theorem that there exist only a finite number of intermediate fields to F and $F(\alpha)$ where α is algebraic over F is now easily completed. Let F have q elements and E be a finite extension of F, $[E:F] = n$. Then E has q^n elements, viz. if $\omega_1, \omega_2, \ldots, \omega_n$ is a basis for E/F, then each $\theta \varepsilon E$ can be written uniquely, $\theta = x_1\omega_1 + x_2\omega_2 + \ldots + x_n\omega_n$, $x_i \varepsilon F$. The $q^n - 1$ non-zero elements in E form a cyclic group under multiplication, i.e. there exists an element γ in E whose powers yield all of the elements of E except zero. Clearly, $E = F(\gamma)$.

CHAPTER V

GALOIS THEORY

5.1 Automorphisms. If a field F is isomorphic to a field $\bar{F}$, $F \simeq \bar{F}$, then there is a mapping $\sigma : F \rightarrow \bar{F}$, given by $\sigma(a) = \bar{a}$, which is one-one, onto, and a homomorphism. Moreover, we have: $\overline{a + b} = \bar{a} + \bar{b}$, $\overline{ab} = \bar{a}\,\bar{b}$, $\bar{1}$ = unit element of $\bar{F}$, $\overline{a^{-1}} = \bar{a}^{-1}$, and $\overline{a/b} = \overline{ab^{-1}} = \bar{a}\,\overline{b^{-1}} = \bar{a}/\bar{b}$. If a mapping σ of F into $\bar{F}$ is known to be a homomorphism, it may be that the image consists of zero alone; otherwise, (Sec. 2.9) the mapping is one-one. If one wishes to show that σ is an isomorphism, one must prove, therefore, that not everything is mapped onto zero and that the mapping is an onto mapping. In the very important special case when $\bar{F} = F$, the isomorphism is called an automorphism. Thus, an automorphism of a field is an isomorphism of the field onto itself. The identity mapping on any field is an automorphism. For the field of complex numbers, the mapping given by $a + bi \rightarrow a - bi$ can easily be shown to be an automorphism. Contrary to a conjecture of Dedekind, there are other automorphisms of the field of complex numbers; indeed, there are as many as there are functions, but Zorn's lemma is needed to describe them.

A symmetry of a point set in space, i.e. a mapping of a point set onto itself preserving distance, provides an intuitive analogue in geometry of an automorphism of a field. In the geometrical analogue, one is interested in the axis of symmetry, i.e. the subset left invariant by the symmetry. For example: for a rotation of an equilateral triangle about its center, the center is the axis; for a reflection in space, the axis is a plane; for the identity transformation, the whole space is the axis.

Although the inner structure of a field that is preserved by an automorphism is more delicate than distance in space, some sort of pseudo-geometric feeling for automorphisms may be conveyed by continuing the analogy. Let the *axis* of an automorphism of a field be the set of elements of the field left undisturbed by the automorphism, e.g. in the automorphism of the complex number field given by $a + bi \rightarrow a - bi$, the set of real numbers is the axis. In general, the axis will contain: 1, the unit element of the field; $1 + 1$ (called 2 even if the field is not a number field); all elements obtainable from 1 by rational operations, e.g. positive integers (pseudo-integers), zero, negative integers, and quotients. Should the field contain the rational numbers, these must be in the axis of any automorphism. In particular, if the field consists of exactly the rational numbers, the only automorphism is the identity.

Example A: Let $F = Q(\sqrt{2})$. Clearly, $[F:Q] = 2$, and each element of F can be uniquely represented by $a + b\sqrt{2}$, a and $b \varepsilon Q$. To find all of the automorphisms of this field, consider $\overline{a + b\sqrt{2}} = \bar{a} + \bar{b}\,\overline{\sqrt{2}} = a + b\,\overline{\sqrt{2}}$. Hence, any automorphism is completely characterized by the image of $\sqrt{2}$. Since $(\sqrt{2})^2 - 2 = 0$, $0 = \bar{0} = (\overline{\sqrt{2}})^2 - 2$, and consequently $\overline{\sqrt{2}} = \pm\sqrt{2}$. Therefore, besides the identity, the only possible automorphism is given by $a + b\sqrt{2} \rightarrow a - b\sqrt{2}$. As an exercise, show that this is an automorphism. Observe that similar reasoning with $\sqrt[3]{2}$ gives $(\overline{\sqrt[3]{2}})^3 - 2 = 0$; but, of the three possibilities for the image of $\sqrt[3]{2}$, the two complex ones are not even in the field considered; hence, the only automorphism of the field $Q(\sqrt[3]{2})$ is the identity.

Example B: Let $F = Q(\sqrt{2}, \sqrt{3})$. For each $\theta \varepsilon F$, $\theta = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$, uniquely, here $\sqrt{6}$ is defined as $\sqrt{2}\,\sqrt{3}$. $\bar{\theta} = a + b\overline{\sqrt{2}} + c\overline{\sqrt{3}} + d\overline{\sqrt{6}}$,

and as in the previous example $\overline{\sqrt{2}} = \pm\sqrt{2}$. Similarly, $\overline{\sqrt{3}} = \pm\sqrt{3}$ and $\overline{\sqrt{6}} = \pm\sqrt{6}$. The eight possibilities apparent are immediately reduced to four since $\overline{\sqrt{6}} = \overline{\sqrt{2}\sqrt{3}} = \overline{\sqrt{2}}\,\overline{\sqrt{3}}$. It can be shown that each of the four remaining cases gives an automorphism. Verify this for one of the three non-trivial cases, as an exercise.

Exercise: Show that the only automorphism for the field R of real numbers is the identity. First, assume an automorphism of R to be a continuous mapping. Secondly, prove it without the assumption of continuity.

This exercise was first proved by the geometer von Staudt; Dedekind, however, was the first to give a modern formulation of it. Von Staudt was interested in the mappings of the projective plane on itself that preserve straight lines.

5.2 Splitting Fields. Let F be a ground field and let $F[x]$ be the polynomial ring over F. Although an algebraic closure of F (Sec. 2.9) exists and is an extension field in which every polynomial of $F[x]$ has a zero, we shall give a method due to Kronecker for constructing a less formidable extension field of F in which a given irreducible polynomial $p(x) \varepsilon F[x]$ has a zero. Consider the canonical mapping $F[x] \longrightarrow F[x]/p(x)\,F[x]$ where $f(x) \rightarrow f(x) + p(x)\cdot\mathcal{O}$ ($\mathcal{O} = F[x]$). Since $F[x]$ is a principal ideal ring, $p(x)\cdot\mathcal{O}$ is a prime ideal and even maximal. Therefore, $F[x] \,/\, p(x)\,F[x]$ is a field. Let $a + p(x)\cdot\mathcal{O}$, the image of the constant a, be designated by $\bar{a}$. The restriction of the map to F is an isomorphism into as the image of 1 is not $p(x)\mathcal{O}$. Let the image of x be called α, i.e. $\alpha = x + p(x)\mathcal{O}$. If $p(x) = x^n + a_1x^{n-1} + \dots + a_n$, then the image of $p(x)$ on the one hand is 0 and on the other is $\alpha^n + \bar{a}_1\alpha^{n-1} + \dots + \bar{a}_n$. If the elements of F replace their images,

the residue class field becomes an extension field of F containing the zero α of the polynomial p(x).

In Galois theory uniqueness of the above extension up to isomorphism is needed. Let $\sigma : F \to \bar{F}$, where $\sigma(a) = \bar{a}$, be an isomorphism of the fields F and $\bar{F}$. Denote by $\bar{p}(x)$ the image of p(x) in the extension of the isomorphism (Sec. 3.2) to the polynomial rings over F and $\bar{F}$. Let $E = F(\alpha)$ be an extension field of F where α is a zero of the polynomial p(x) irreducible in $F[x]$ and of degree n. Similarly, let $\bar{E} = \bar{F}(\bar{\alpha})$ where $\bar{\alpha}$ is a zero of $\bar{p}(x)$ irreducible in $\bar{F}[x]$. Contention: σ can be extended to an isomorphism τ of $F(\alpha)$ and $\bar{F}(\bar{\alpha})$ such that $\bar{\alpha}$ is the image of α, i.e. τ restricted to F is σ and $\tau(\alpha) = \bar{\alpha}$. Proof: Any element $\theta \in F(\alpha)$ is of the form $\theta = f(\alpha) = a_o + a_1\alpha + \ldots + a_r\alpha^r$ where it will be advantageous not to restrict r even though $r \leq n - 1$ would be sufficient. This means, however, that uniqueness is being sacrificed, e.g. 0 and $p(\alpha)$ represent the same element. Define $\tau(\theta) = \bar{f}(\bar{\alpha}) = \bar{a}_o + \bar{a}_1\bar{\alpha} + \ldots + \bar{a}_r\bar{\alpha}^r$. We have to show that $\tau(\theta)$ is well-defined: $f(\alpha) = g(\alpha) \Longrightarrow f(\alpha) - g(\alpha) = 0 \Longrightarrow p(x) \mid f(x) - g(x) \Longrightarrow f(x) - g(x) = p(x)\,q(x) \Longrightarrow \bar{f}(x) - \bar{g}(x) = \bar{p}(x)\,\bar{q}(x) \Longrightarrow \bar{p}(x) \mid \bar{f}(x) - \bar{g}(x) \Longrightarrow \bar{f}(\bar{\alpha}) - \bar{g}(\bar{\alpha}) = 0 \Longrightarrow \bar{f}(\bar{\alpha}) = \bar{g}(\bar{\alpha})$. Since it is obvious that τ is a homomorphism, τ is the extended isomorphism desired.

If F is a ground field and if an arbitrary polynomial $f(x) \in F[x]$ where f(x) is of degree n and not necessarily irreducible, then there exist extension fields of F in which f(x) splits into linear factors, i.e. in which $f(x) = c(x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n)$. For instance, the algebraic closure of F may be used. Also an extension of F to a field where one of the non-linear factors of f(x), irreducible in $F[x]$, has a

zero can be constructed. This process can be repeated as often as necessary with one of the irreducible non-linear factors remaining in the new field until all the factors are linear. The degree of the final extension over F will be at most $n!$ as the first extension has degree at most n, the second at most $n - 1$, etc. By the splitting field of $f(x)$ is meant the smallest extension field of F yielding only linear irreducible factors. The uniqueness up to isomorphism of the splitting field of $f(x)$ will follow from the more general result of the next paragraph.

As before, let $\sigma : F \to \overline{F}$ be an isomorphism of the fields F and $\overline{F}$, and let $\overline{f}(x)$ be the image of $f(x)$ in the extension of the isomorphism to the polynomial rings. Further, let E and $\overline{E}$ be the splitting fields for $f(x)$ and $\overline{f}(x)$ respectively. Contention: σ can be extended to an isomorphism τ of the splitting fields E and $\overline{E}$ such that τ restricted to F is σ. Proof: In E, we have $f(x) = c(x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n)$. In F, $f(x) = c(x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_s)p_1(x)\, p_2(x) \dots p_k(x)$ where each $p_i(x)$ is irreducible and has degree greater than 1. Induction on m, the number of roots not in F, will be used $(m = n - s)$.

1) $m = 0$. In this case all roots are in F. Hence, $E = F$ and $\overline{E} = \overline{F}$ since $\overline{f}(x) = \overline{c}(x - \overline{\alpha}_1)(x - \overline{\alpha}_2) \dots (x - \overline{\alpha}_n)$ in this case.

2) $m > 0$. Say α_i not in F, then α_i must be a root of say $p_1(x)$. In $\overline{F}$, we have $\overline{f}(x) = \overline{c}(x - \overline{\alpha}_1)(x - \overline{\alpha}_2) \dots (x - \overline{\alpha}_s)\, \overline{p}_1(x)\, \overline{p}_2(x) \dots \overline{p}_k(x)$. Since $\overline{E}$ is a splitting field, $\overline{p}_1(x)$ in particular must split in $\overline{E}$, i.e. $\overline{p}_1(x)$ has a certain root $\overline{\alpha}_i$ in E. By the previous uniqueness proof, σ can be extended to an isomorphism σ' of the fields $F(\alpha_i)$ and $\overline{F}(\overline{\alpha}_i)$. Consider $F(\alpha_i)$ and $\overline{F}(\overline{\alpha}_i)$ as ground fields with E and $\overline{E}$ respectively as splitting fields. As the number of roots of $f(x)$ not in $F(\alpha_i)$ is less than m, the inductive assumption allows the extension of σ', and consequently σ, to an isomorphism τ between E and $\overline{E}$.

$$\begin{tikzcd} & K & \\ E \arrow[ur, dash] \arrow[rr, "\sigma"] & & E' \arrow[ul, dash] \\ & F \arrow[ul, dash] \arrow[ur, dash] & \end{tikzcd}$$

Fig. A

In Galois theory one frequently encounters a situation where four fields F, E, E', and K are such that $F \subset E \subset K$, $F \subset E' \subset K$, and an isomorphism $\sigma : E \rightarrow E'$ is given where σ restricted to F is the identity mapping; in such a case σ is called an isomorphism relative to F or a relative isomorphism of E/F. Let α be an algebraic element of E over F, $\sigma(\alpha) = \beta$, and $p(x) = \text{Irr}(\alpha, F) = x^n + a_1x^{n-1} + \dots + a_n$ $(a_i \in F)$. Applying σ to $0 = \alpha^n + a_1\alpha^{n-1} + \dots + a_n$ gives $0 = \beta^n + a_1\beta^{n-1} + \dots + a_n$. Hence, $\sigma(\alpha) = \beta$ must also be a zero of p(x). Since E and E' are in the common field K, p(x) has a limited number of zeros; consequently, there are only a finite number of possibilities for E' in case $E = F(\alpha)$. More generally, the number of relative isomorphisms of E/F is finite when E is a finite extension of F. In this case $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$, and every $\theta \in E$ can be expressed as a polynomial in $\alpha_1, \alpha_2, \dots, \alpha_n$ (cf. p. 47), i.e. $\theta = \phi(\alpha_1, \alpha_2, \dots, \alpha_n)$, a polynomial in $\alpha_1, \alpha_2, \dots, \alpha_n$. By the rules for an isomorphism, $\sigma(\theta) = \phi\big(\sigma(\alpha_1), \sigma(\alpha_2), \dots, \sigma(\alpha_n)\big)$; in other words, to describe σ, it is sufficient to know all $\sigma(\alpha_i)$. Each $\sigma(\alpha_i)$ must be a zero of any polynomial having α_i as a zero. Hence, there are but a finite number of possibilities for each $\sigma(\alpha_i)$, and the number of relative isomorphisms is certainly finite.

Let E/F be the splitting field of a polynomial $f(x) \in F[x]$; hence, $f(x) = c(x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n)$ in E and $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. Let E be embedded in a field K, and consider the relative isomorphisms of E/F. If σ is a relative isomorphism of E/F, then each $\sigma(\alpha_i)$ must be a zero of f(x) just as for p(x) previously (irreducibility was not used). The extension of σ to an isomorphism of the polynomial rings may be

applied to $f(x)$ to obtain $f(x) = c\bigl(x - \sigma(\alpha_1)\bigr)\bigl(x - \sigma(\alpha_2)\bigr) \ldots \bigl(x - \sigma(\alpha_n)\bigr)$. From the uniqueness of factorization of the polynomial, it follows that σ permutes the generators α_i, and $\sigma : E \xrightarrow{\text{onto}} E$ as $\sigma(E) = F\bigl(\sigma(\alpha_1), \sigma(\alpha_2), \ldots, \sigma(\alpha_n)\bigr) = E$. In other words, σ is a relative automorphism, and the common comprehending field K is no longer needed. A counter example when E is not a splitting field can be constructed by letting $F = Q$, $E = Q(\sqrt[3]{2})$, $K = C$, and $E' = Q(\theta)$ where $\theta = \omega\sqrt[3]{2}$, $\omega^2 + \omega + 1 = 0$. E and E' are isomorphic but not equal for E' contains complex numbers whereas E contains only real numbers.

$$\begin{tikzcd}
 & C \arrow[dl, dash] \arrow[dr, dash] & \\
Q(\sqrt[3]{2}) \arrow[rr] & & Q(\theta) \\
 & Q \arrow[ul, dash] \arrow[ur, dash] &
\end{tikzcd}$$

Fig. B

$$\begin{tikzcd}
 & K \arrow[dl, dash] \arrow[dr, dash] & \\
E \arrow[rr, "\tau"] \arrow[d, dash] & & E_1(E) \arrow[d, dash] \\
E_0 \arrow[rr, "\sigma"] & & E_1 \\
 & F \arrow[ul, dash] \arrow[ur, dash] &
\end{tikzcd}$$

Fig. C

Let E_0 be an intermediate field to F and E, $F \subset E_0 \subset E$, let $f(x) \in F[x]$, and let σ be a relative isomorphism, $\sigma : E_0 \to E_1$, in a field K containing E and E_1. View E as the splitting field of our old polynomial $f(x) \in F[x] \subset E_0[x]$ where $f(x) = c(x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n)$ in E; similarly, view $E_1(E) = E_1(\alpha_1, \alpha_2, \ldots, \alpha_n)$, intermediate to E_1 and K, as the splitting field of the image in $E_1[x]$ of $f(x)$ under σ extended to the polynomials; the image of $f(x)$ is also $f(x)$ in this case. Consequently, by an earlier result of this section, there exists an isomorphism $\tau : E \xrightarrow{\text{onto}} E_1(E)$ such that τ restricted to E_0 is σ. τ is a relative isomorphism of E/F inside the comprehending field K. Hence, by the result of the previous paragraph, τ is a relative automorphism, $E_1(E) = E$, and in particular $E_1 \subset E$. Therefore, any relative isomorphism of a field

intermediate to F and a splitting field E can be extended to a relative automorphism of E/F; in other words, the relative isomorphisms of the intermediate fields can be viewed as restrictions of the relative automorphisms of E/F. For an example to show that this result is false when E is not a splitting field, let $K = C$, $E = Q(\sqrt[4]{2})$, $E_o = Q(\sqrt{2})$, and $F = Q$. Viewing E embedded in C, four isomorphisms of E relative to F are possible, viz. those characterized by having as image of $\sqrt[4]{2}$, $\pm\sqrt[4]{2}$ or $\pm i\sqrt[4]{2}$. The last two possibilities are immediately eliminated as automorphisms since E contains only real numbers. In each of the two remaining cases $\sqrt{2} = (\pm\sqrt[4]{2})^2 \longrightarrow \sqrt{2}$. Hence, the isomorphism of the intermediate field E_o onto E characterized by $\sqrt{2} \longrightarrow -\sqrt{2}$ ($E_1 = E_o$ in this case) is not obtainable by restricting an automorphism of E to E_o.

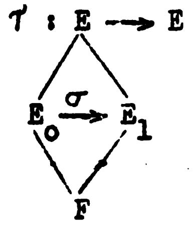


Fig. C (condensed)

Let E/F be a splitting field, $\beta \varepsilon E$, and $p(x) = \text{Irr}(\beta, F)$. Further, take K as the splitting field of p(x) over E, $F(\beta)$ as E_o, and $F(\gamma)$ as E_1 where γ is another zero of p(x). As β and γ are zeros of the same irreducible polynomial, there is an isomorphism $\sigma : E_o \longrightarrow E_1$. This means by virtue of the preceding paragraph $E_1 \subset E$ and especially $\gamma \varepsilon E$. Hence, p(x) splits in E. Consequently, if p(x) is an irreducible polynomial having one zero β in a splitting field E/F, then p(x) splits completely in E; moreover, there exists an automorphism of E/F that moves one zero of p(x) into any other.

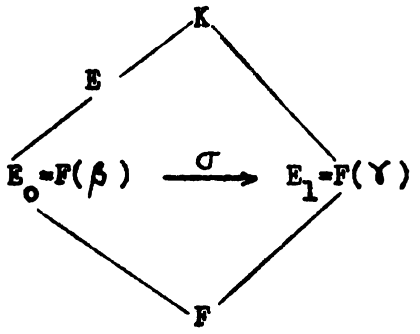


Fig. D

Conversely, if E is a finite extension of F, i.e. $F(\alpha_1, \alpha_2, \ldots, \alpha_s)$, and if every irreducible polynomial $p(x) \varepsilon F[x]$ with one root in E splits into linear factors in E, then E/F is a splitting field.

Proof: Let $p_i(x) = Irr(\alpha_i, F)$ so that each $p_i(x)$ has a root α_i in E. By hypothesis, then, $p_i(x)$ splits in E. Hence, E is the splitting field of $f(x) = \prod p_i(x)$ since the roots of $f(x)$ lie in E and generate E. If the hypothesis of this theorem is satisfied, the extension field E is said to be automorphic over F : more briefly, E/F is automorphic. Although normal is sometimes used instead of automorphic, normal will later be used in a more restricted situation.

Let E/F be automorphic, and let G be the set (finite) of all relative automorphisms of E/F. After the group structure of G is given, it will be called the Galois group of this extension E/F. It has been shown that any relative isomorphism of an intermediate field can be extended to a relative automorphism of E/F and that any element of E can only be moved into a root of the irreducible polynomial in F[x] which this element satisfies and can really be so moved by a relative automorphism of E/F. Let K be the set of elements of E left fixed by all of G, i.e. $K = \{\alpha \mid \alpha \varepsilon E, \sigma(\alpha) = \alpha$ for all $\sigma \varepsilon G\}$. It is easy to show that the set of elements left fixed under any set of automorphisms is a field; in particular, K is a field. By the definition of a relative automorphism $F \subset K$. In most cases, but not always, K = F. If α is an element of K and $p(x) = Irr(\alpha, F)$, then p(x) splits in E. None of its roots can be different from α since, otherwise, there would be an automorphism moving α so that α would not lie in K. Hence, $p(x) = (x - \alpha)^n$ in E. Conversely, this assures us that α lies in K. Hence, K may be characterized as the

set of all $\alpha \varepsilon E$ where $p(x) = Irr(\alpha, F)$ has only the root α. An element α with this property is called a <u>totally inseparable</u> element; a <u>separable</u> element is defined as one for which $p(x)$ has no $\overset{\text{(duplicated)}}{\text{multiple}}$ roots. The existence of totally inseparable elements will be postponed momentarily for a definition. The <u>characteristic</u> of a field F is a property of its additive group; it is defined as the period of its non-zero elements under addition. If no "power" of a non-zero element is the unit element of the additive group, the period, and consequently the characteristic, is defined to be 0! In the homomorphism of the additive group of integers $Z \longrightarrow G$ where $n \longrightarrow n \cdot a = (\underbrace{a + a + \ldots + a}_{n \text{ terms}})$, the kernel consists of either 0 alone in which case the characteristic is 0 or multiples of d. In the latter case, all non-zero elements have the same period since $n \cdot a = n \cdot b(b^{-1} a)$ means $n \cdot b = 0 \Longrightarrow n \cdot a = 0$. Consequently, it suffices to examine the period of 1. Consider $n \cdot 1 = 1 + 1 + \ldots + 1 = 0$ and n the smallest such positive integer if any such n exist. If n exists, it must be a prime; $\overset{\text{otherwise}}{\text{e.g.}}$, $6 \cdot 1 = (3 \cdot 1)(2 \cdot 1) = (1 + 1 + 1)(1 + 1) = 0 \Longrightarrow 3 \cdot 1 = 0$ or $2 \cdot 1 = 0$. Moreover, the field Z/pZ is an example of a field with given prime characteristic p. For F of characteristic p: $(a \pm b)^p = a^p \pm b^p$ since $p \mid \binom{p}{i}$ for $1 \leq i \leq p - 1$; more generally, $(a \pm b)^{p^n} = a^{p^n} \pm b^{p^n}$; extraction of p^n-th roots is unique as $a^{p^n} = b^{p^n} \Longrightarrow a^{p^n} - b^{p^n} = 0 \Longrightarrow (a - b)^{p^n} = 0 \Longrightarrow a = b$. It should be observed that the characteristic of an extension field is always the same as the characteristic of the ground field. In returning to the question of existence of a totally inseparable element α with

$p(x) = \mathrm{Irr}(\alpha, F) = (x - \alpha)^n \varepsilon F[x]$, we shall distinguish between two cases.

1) Characteristic of $F = 0$. The second coefficient in the expansion of $p(x)$ is $-n\alpha$ which must be in F. Hence, $n\alpha \varepsilon F$ where $n = 1 + 1 + \ldots + 1 \ \varepsilon F$ and $n \neq 0$. Therefore, $\alpha \ \varepsilon F$. This means $n = 1$. In this case there are no totally inseparable elements outside of F and the fixed field under G is F.

2) Characteristic of $F = p$, a prime. Let $n = p^r \cdot m$, $p \nmid m$. $p(x) = \left[(x - \alpha)^{p^r}\right]^m = \left(x^{p^r} - \alpha^{p^r}\right)^m$, and the negative of the second coefficient is $m\alpha^{p^r} \ \varepsilon F$. Division by $m \neq 0$ gives α^{p^r} = certain $a \ \varepsilon F$. Consequently, α must be a root of $x^{p^r} - a \ \varepsilon F[x]$, but the degree of $p(x) = n = p^r \cdot m \leq p^r$. Hence, $m = 1$, and the only possibility for α is a p^r-th root of an element in F. Conversely, it will be shown that any p^r-th root of any element in the ground field F is totally inseparable.

Let $a \ \varepsilon F$. Then, $\alpha = \sqrt[p^r]{a} \Longrightarrow \alpha^{p^r} = a \Longrightarrow \alpha$ a root of $x^{p^r} - a$, not necessarily irreducible. However, the irreducible polynomial having α as a root will divide $x^{p^r} - a$, which has been shown to have all roots equal. Therefore, the irreducible polynomial having α as a root also has all roots equal, and α is totally inseparable. In case the characteristic of F is the prime p, the fixed field under G consists of those p^r-th roots of the elements of F that are in E.

Assume F to be a ground field, $f(x) \in F[x]$, E the splitting field of $f(x)$, G the set of all relative automorphisms of E/F, and K the fixed field under G. That each irreducible factor of $f(x)$ has only simple roots in E (in which case $f(x)$ is called separable) will now be shown to be a sufficient condition for $K = F$. Later, this condition will also turn out to be necessary. Assume $f(x)$ separable, i.e. $p(x)|f(x)$, $p(x)$ irreducible $\Longrightarrow p(x) = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_r)$ in E and $\alpha_1, \alpha_2, \ldots, \alpha_r$ are distinct. If $f(x) = p_1(x)^{k_1} p_2(x)^{k_2} \ldots p_s(x)^{k_s}$ where $k_i \geq 1$ and the $p_i(x)$ are distinct and irreducible, then $f(x)$ can be replaced by the polynomial $g(x) = p_1(x)\, p_2(x) \ldots p_s(x)$, since $f(x)$ splits in E if and only if $g(x)$ splits in E. Furthermore, for $i \neq j$, $p_i(x)$ and $p_j(x)$ will not have a factor in common as the common root α determines uniquely $p(x) = \mathrm{Irr}\,(\alpha, F)$. Therefore, without loss of generality, $f(x)$ will be assumed to have only simple roots. Contention: $K = F$. The method of proof will be induction on m, the number of roots in the splitting field of $f(x)$ which are not in F.

1) $m = 0$. No roots are outside of F, and $f(x)$ splits in F in this case. Consequently, $E = F$ which implies $K = F$.

2) $m > 0$. Say $\alpha_1 \notin F$ and α_1 a root of say $p_1(x)$ where $p_1(x)|f(x)$ and $p_1(x)$ is irreducible. Just as E is the splitting field over F of $f(x) \in F[x]$, E is also the splitting field over $F(\alpha_1)$ of the same $f(x)$ viewed this time as an element of $F(\alpha_1)[x]$. In the latter case there is one more root in the ground field, and consequently the corresponding value of m is smaller. The set $\bar{G}$ of all relative automorphisms of E over $F(\alpha_1)$ is a subset of those of E over F, i.e. $\bar{G} \subset G$. This means that any

$$\begin{tikzcd}
E \arrow[d, dash] \\
F(\alpha_1) \arrow[d, dash] \\
F
\end{tikzcd}$$

Fig. E

element left fixed by G is especially left fixed by $\bar{G}$; therefore, $\bar{K} \supset K$ where $\bar{K}$ is the field left fixed under $\bar{G}$. As f(x) is still separable in the new situation, we are now in a position to use the inductive assumption, which yields $\bar{K} = F(\alpha_1)$, $F(\alpha_1)$ being the ground field. It follows that $K \subset F(\alpha_1)$. Let $p_1(x) = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_r)$ where the α_i are distinct and the degree of $p_1(x) = r = [F(\alpha_1):F]$. Consider the situation illustrated by Fig. G. The isomorphism of $F(\alpha_1)$ onto $F(\alpha_2)$, which exists because α_1 and α_2 are roots of the same irreducible polynomial, leaves F fixed and moves α_1 into α_2. So there exists an automorphism of E which produces this map. Call σ_i an element of G which maps α_1 onto α_i, $i \leq r$. In Fig. F observe that $F(\alpha_1) = K(\alpha_1)$, and let $q(x) = \mathrm{Irr}(\alpha_1, K)$. q(x) has coefficients in K and $q(\alpha_1) = 0$. Now apply each σ_i to $q(\alpha_1) = 0$ to obtain $q(\alpha_i) = 0$ since the coefficients belong to K and therefore are left fixed. As $\alpha_1, \alpha_2, \ldots, \alpha_r$ are distinct, q(x) has at least r distinct roots; hence, $r \leq$ degree of $q(x) = [F(\alpha_1):K]$. It follows that $r = [F(\alpha_1):F] = [F(\alpha_1):K][K:F] \geq r[K:F]$. Consequently, $[K:F] = 1$ and $K = F$.

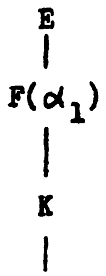


Fig. F

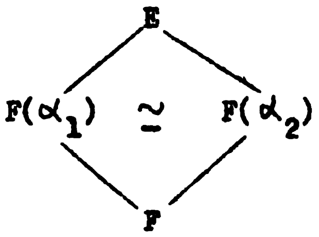


Fig. G

5.3 **Fixed Fields under Automorphisms.** At this stage we take another viewpoint; we start with a set of automorphisms of any field and study ~~their~~ its structure. Let E be any field, and let $\sigma_1, \sigma_2, \ldots, \sigma_n$ be n distinct automorphisms of E, which should be viewed as functions $\sigma_1(x), \sigma_2(x), \ldots, \sigma_n(x)$ where x ranges over E. By the nature of an automorphism, the function values also range over E. We shall only prove that these automorphisms are linearly unrelated, nevertheless they are also totally unrelated algebraically except when E is finite. Contention: The functions $\sigma_1(x), \sigma_2(x), \ldots, \sigma_n(x)$ are linearly independent over E, i.e. if $c_1, c_2, \ldots, c_n \in E$ and $c_1\sigma_1(x) + c_2\sigma_2(x) + \ldots + c_n\sigma_n(x) = 0$ for all $x \in E$, then all $c_i = 0$. Otherwise among the n-tuples in E, $c_1, c_2, \ldots, c_n$, not all zero, for which $c_1\sigma_1(x) + c_2\sigma_2(x) + \ldots + c_n\sigma_n(x)$ for all $x \in E$, let k be the smallest number of non-zero c_i in any such n-tuple. We have then, by renumbering if necessary, $c_1\sigma_1(x) + c_2\sigma_2(x) + \ldots + c_k\sigma_k(x) = 0$ for all $x \in E$ and no c_i is zero. First, $k > 1$ since $c_1\sigma_1(x) = 0$ for all $x \in E \Longrightarrow c_1\sigma_1(1) = 0 \Longrightarrow c_1 = 0$. Furthermore, for any $a \in E$, ax is also an element of E; consequently, $c_1\sigma_1(ax) + c_2\sigma_2(ax) + \ldots + c_k\sigma_k(ax) = 0$. Using $\sigma_i(ax) = \sigma_i(a)\sigma_i(x)$ gives $c_1\sigma_1(a)\sigma_1(x) + c_2\sigma_2(a)\sigma_2(x) + \ldots + c_k\sigma_k(a)\sigma_k(x) = 0$. However, multiplication by $\sigma_k(a)$ in the initial equation involving k yields $c_1\sigma_k(a)\sigma_1(x) + c_2\sigma_k(a)\sigma_2(x) + \ldots + c_k\sigma_k(a)\sigma_k(x) = 0$. By subtraction and the last two equations,

$$c_1\big(\sigma_1(a) - \sigma_k(a)\big)\sigma_1(x) + c_2\big(\sigma_2(a) - \sigma_k(a)\big)\sigma_2(x) + \ldots + c_{k-1}\big(\sigma_{k-1}(a) - \sigma_k(a)\big)\sigma_{k-1}(x) = 0$$

for all $a, x \in E$. In particular, for $a \in E$ such that $\sigma_1(a) \neq \sigma_k(a)$ (this a exists since $\sigma_1 \neq \sigma_k$ as

$k > 1$), we have obtained a shorter equation, contradictory to the choice of k. As an example of proof analysis, i.e. a study of exactly what is used in a proof for the purpose of generalization, observe that the proof just given required the following: $\sigma_i(ax) = \sigma_i(a)\sigma_i(x)$; $\sigma_1(1) \neq 0$; $c_i \in E$; σ-values in E; the arguments of the functions, e.g. a and x, not occurring except as independent variables, need not be elements of E. Verify that the proof goes through if G is a multiplicative group, σ_i is a homomorphism (called a *character* of G) $\sigma_i : G \longrightarrow$ the multiplicative group $E - \{0\}$, and "$\in E$" is appropriately replaced by "$\in G$." The method of proof analysis has been exploited by mathematicians only in the last fifty years.

Corollary: $\sigma_1(x) + \sigma_2(x) + \dots + \sigma_n(x)$ cannot be zero for all $x \in E$.

The theorem of the previous paragraph allows us to study the fixed field. Let E be any field and $\sigma_1, \sigma_2, \dots, \sigma_n$ be n distinct automorphisms of E. Let F be the field left fixed by these automorphisms, i.e. $F = \{\alpha \mid \alpha \in E,\ \sigma_i(\alpha) = \alpha$ for all $i\}$. Contention: $[E:F] \geq n$. Otherwise, we shall show the σ_i to be linearly dependent. Assume $[E:F] = m < n$. Then E has a basis $\omega_1, \omega_2, \dots, \omega_m$ over F, i.e. any $x \in E$ has the unique form $x = a_1\omega_1 + a_2\omega_2 + \dots + a_m\omega_m$ with $a_i \in F$. The idea of the proof is to find a set of c_i not all zero such that $c_1\sigma_1(x) + c_2\sigma_2(x) + \dots + c_n\sigma_n(x) = 0$ for all $x \in E$. By the additive property of automorphisms, we need only to establish this for $a_1\omega_1, a_2\omega_2, \dots, a_m\omega_m \in E$, i.e. taking $x = a_i\omega_i$, we need

$c_1\sigma_1(a_i\omega_i) + c_2\sigma_2(a_i\omega_i) + \dots + c_n\sigma_n(a_i\omega_i) = 0$ for $i = 1, 2, \dots, m$.

As $\sigma_j(a_i\omega_i) = a_i\sigma_j(\omega_i)$, it would suffice to take each $a_i = 1$ and to have a non-trivial solution of the system of m homogeneous linear equations

$c_1\sigma_1(\omega_i) + c_2\sigma_2(\omega_i) + \dots + c_n\sigma_n(\omega_i) = 0$ $(i = 1, 2, \dots, m)$ in the n unknowns $c_1, c_2, \dots, c_n$. Since $m < n$, this is given by a well-known result of vector spaces. For a more formal approach we could say: Consider the system of m homogeneous linear equations $c_1\sigma_1(\omega_i) + c_2\sigma_2(\omega_i) + \dots + c_n\sigma_n(\omega_i) = 0$ $(i = 1, 2, \dots, m)$ in the n unknowns $c_1, c_2, \dots, c_n$. These have, since $m < n$, a non-trivial solution in E. Multiplication of the i-th equation by a_i(equal to any $\sigma_j(a_i)$), $a_i \varepsilon F$, and addition over all i gives $c_1\sigma_1(x) + c_2\sigma_2(x) + \dots + c_n\sigma_n(x) = 0$ for all $x \varepsilon E$. But this contradicts the previous theorem.

<u>Example</u>: Let k be any field and $x_1, x_2, \dots, x_n$ independent variables. Let E be the field of rational functions in the n variables with coefficients from k, i.e. $E = k(x_1, x_2, \dots, x_n)$. E has n! obvious automorphisms, viz. the permutations of x_i. If $f(t) = (t - x_1)(t - x_2) \dots (t - x_n)$, then $f(t) = t^n + a_1 t^{n-1} + \dots + a_n$ where

$$a_1 = -\sum x_i, \quad a_2 = \sum_{i<j} x_i x_j, \quad a_3 = -\sum_{i<j<r} x_i x_j x_r,$$

etc. If F is the fixed field under these n! automorphisms, we have $F_o \subset F \subset E$ where $F_o = k(a_1, a_2, \dots, a_n)$, as the a_i, and consequently the rational functions of the a_i, are in F. Observe that E can certainly be obtained from F_o by adjoining successively $x_1, x_2, \dots, x_n$. Since x_1 is a root of $f(t)$, not necessarily irreducible over F_o, $[F_o(x_1):F_o] \leq n$; since x_2 is a root of $\frac{f(t)}{t - x_1}$, $[F_o(x_1, x_2) : F_o(x_1)] \leq n - 1$; etc. Therefore $[E:F_o] \leq n!$; however (by the foregoing theorem), $[E:F] \geq n!$ and $[F:F_o] \geq 1$ (as $F_o \subset F$) $\Longrightarrow [E:F_o] \geq n!$. Consequently, all equalities must hold, and $F = F_o$. This result is known as the fundamental theorem of symmetric functions.

If τ and σ are automorphisms of a field E, then the composite mapping $\sigma\tau$ where $\sigma\tau(x) = \sigma\big(\tau(x)\big)$ is also an automorphism of E, since $\sigma\tau$ is clearly onto and $\sigma\tau(x \overset{+}{\cdot} y) = \sigma\big(\tau(x \overset{+}{\cdot} y)\big) = \sigma\big(\tau(x) \overset{+}{\cdot} \tau(y)\big) = \sigma\tau(x) \overset{+}{\cdot} \sigma\tau(y)$. The composition of automorphisms, as for maps in general, is associative. The identity mapping on E serves as the unit element from both sides for the automorphisms of E, and the inverse automorphism of a given automorphism σ is simply the inverse mapping σ^{-1} (Sec. 1.3). Hence, the set of all automorphisms of a field E forms a group. Consequently, for an arbitrary set of automorphisms of E to form a subgroup, closure and the existence of an inverse within the set will suffice. We have seen that a field E and a finite set $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ of distinct automorphisms of E determine a fixed field F such that $[E:F] \geq n$. If this set of automorphisms is not a group, then say $\sigma_1\sigma_2 = \sigma_{n+1}$ is not in the set. The fixed field of E under $\{\sigma_1, \sigma_2, \ldots, \sigma_{n+1}\}$, on the one hand, is contained in F as it is in particular fixed under $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$; on the other hand, it contains F since F itself is fixed under these n + 1 automorphisms. Therefore, the fixed field under these n + 1 automorphisms is also F, and $[E:F] \geq n + 1 > n$. By adding to the set $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, products and inverses of the σ_i a group of automorphisms may be constructed, but it may happen that this group is infinite, i.e. that $[E:F] = \infty$.

<u>Exercise</u>: Take E = k(x) to be the field of rational functions f(x) of a single indeterminate x where k is a field of characteristic 0. Let the automorphism σ of E be given by $f(x) \longrightarrow f(x + 1)$. Under the set of automorphisms $\{\sigma^n\}$ where n is an integer and σ^n is given by $f(x) \longrightarrow f(x + n)$, show that the fixed field is k. Also determine the fixed field in case k has characteristic $p > 0$.

If $G = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ is a finite group of automorphisms of E, then $[E:F] \leq n$ (hence $= n$). In the proof we shall show that more than n elements of E are linearly dependent over F. Let $\alpha_1, \alpha_2, \ldots, \alpha_m \in E$ and $m > n$. A non-trivial solution $x_1, x_2, \ldots, x_m$ in F of $\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m = 0$ is needed. As one of the σ_i must be the identity, this equation is contained in the system of homogeneous linear equations in E whose matrix is

$$\begin{pmatrix} \sigma_1(\alpha_1) & \sigma_1(\alpha_2) & \ldots & \sigma_1(\alpha_m) \\ \sigma_2(\alpha_1) & \sigma_2(\alpha_2) & \ldots & \sigma_2(\alpha_m) \\ & \ldots & & \\ \sigma_n(\alpha_1) & \sigma_n(\alpha_2) & \ldots & \sigma_n(\alpha_m) \end{pmatrix} ,$$

i.e. whose i-th equation is $\sigma_i(\alpha_1)x_1 + \sigma_i(\alpha_2)x_2 + \ldots + \sigma_i(\alpha_m)x_m = 0$. A non-trivial solution in F of this system will suffice. Consisting of more unknowns (m) than equations (n), this system has non-trivial solutions in E. Applying $\tau \in G$ to the system gives a new system of homogeneous linear equations in the unknowns $\tau(x_1), \tau(x_2), \ldots, \tau(x_m)$ with matrix

$$\begin{pmatrix} \tau\sigma_1(\alpha_1) & \tau\sigma_1(\alpha_2) & \ldots & \tau\sigma_1(\alpha_m) \\ \tau\sigma_2(\alpha_1) & \tau\sigma_2(\alpha_2) & \ldots & \tau\sigma_2(\ _m) \\ & \ldots & & \\ \tau\sigma_n(\alpha_1) & \tau\sigma_n(\alpha_2) & \ldots & \tau\sigma_n(\alpha_m) \end{pmatrix} .$$

Because of the group property, this matrix differs from the original matrix only in the arrangement of its rows, i.e. the equations of the original system have been permuted. Hence, if $x_1, x_2, \ldots, x_m$ is a solution of the

original system $\tau(x_1)$, $\tau(x_2)$, ..., $\tau(x_m)$ is also a solution, i.e. for each i, $\sigma_i(x_1)$, $\sigma_i(x_2)$, ..., $\sigma_i(x_m)$ is a solution. Define the *trace* of x, $S(x) = \sigma_1(x) + \sigma_2(x) + \ldots + \sigma_n(x)$. Then, $S(x_1)$, $S(x_2)$, ..., $S(x_m)$ is also a solution, since the sum of solutions of a system of homogeneous linear equations gives another solution. Moreover, this last solution is *in F* for $\tau S(x) = \tau\bigl(\sigma_1(x) + \sigma_2(x) + \ldots + \sigma_n(x)\bigr) =$ $\tau\sigma_1(x) + \tau\sigma_2(x) + \ldots + \tau\sigma_n(x) =$ (by the group property) $\sigma_1(x) + \sigma_2(x) + \ldots + \sigma_n(x) = S(x)$. To be certain that a non-trivial solution in F can be obtained in this manner, let x_1, x_2, ..., x_m be a non-trivial solution in E. Since we have a homogeneous linear system, cx_1, cx_2, ..., cx_m is also a solution in E, and $S(cx_1)$, $S(cx_2)$, ..., $S(cx_m)$ is a solution in F. Say $x_1 \neq 0$, and select c in such a way that cx_1 is an element of E with $S(cx_1) \neq 0$. This can be done for, by the previous corollary, S(x) cannot be zero for all x.

Let E be a field and F the fixed field under E for the group of automorphisms $G = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$. Each σ_i is a relative automorphism of E/F. Since $[E:F] = n$, G consists of all the relative automorphisms of E/F. Otherwise τ is another one, and the fixed field under E for the automorphisms σ_1, σ_2, ..., σ_n and τ is still F; but this would mean $n = [E:F] \geq n + 1$. Thus, G can be constructed from F by taking all the relative automorphisms of E/F. As E over F is finite, the elements of E are algebraic. Looking more closely, let us take $\alpha \in E$ and try to construct $Irr(\alpha, F)$. Let $\alpha_1 (= \alpha)$, α_2, ..., α_r be the distinct images of α where $r \leq n$, e.g. α may even be in F. Consider $p(x) =$ $(x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_r)$. First, we shall show that p(x) has coefficients in F. Let $\tau \in G$ and apply τ to $p(x)$: $\tau\bigl(p(x)\bigr) =$ $\bigl(x - \tau(\alpha_1)\bigr)\bigl(x - \tau(\alpha_2)\bigr) \ldots \bigl(x - \tau(\alpha_r)\bigr)$. As α_1, α_2, ..., α_r can

be obtained by deleting the repetitions from $\sigma_1(\alpha)$, $\sigma_2(\alpha)$, ..., $\sigma_n(\alpha)$ or from $\tau\sigma_1(\alpha)$, $\tau\sigma_2(\alpha)$, ..., $\tau\sigma_n(\alpha)$, it follows that $\tau(\alpha_1)$, $\tau(\alpha_2)$, ..., $\tau(\alpha_r)$ is another arrangement of $\alpha_1, \alpha_2, \ldots, \alpha_r$. Therefore, $\tau\big(p(x)\big) = p(x)$ and the coefficients of p(x) must be left fixed under G. Secondly, p(x) is the polynomial in F of lowest degree that α can satisfy. Assume $f(x) \varepsilon F[x]$ and $f(\alpha) = 0$. Applying σ_i to this equation yields $f\big(\sigma_i(\alpha)\big) = 0$. This means that f(x) has $\alpha_1, \alpha_2, \ldots, \alpha_r$ as roots, and consequently $p(x)|f(x)$. As a consequence, E/F is automorphic and separable. E/F is automorphic since E is a finite extension field over F and an irreducible polynomial having a root α in E splits in E. The irreducible polynomial must be the p(x) described above. E/F is called *separable* since each of its elements is separable with respect to F (cf. p. 64). When E/F is both automorphic and separable, it is said to be *normal*.

Let E be any field and Γ be the group of all automorphisms of E. To each finite subgroup G of Γ there corresponds a fixed field F. It has just been shown that the subfield F is normal under E. Moreover, we shall show that all subfields F normal under E are obtainable in this manner. Let F be a subfield such that E is the splitting field of a separable polynomial $f(x) \varepsilon F[x]$. Then F is the fixed field under the set of all relative automorphisms of E/F, which in fact is a finite group. Hence, the correspondence between the finite subgroups G of Γ and the subfields F normal under E is one to one. This result is known as the *fundamental theorem of Galois theory*. As a corollary, the splitting field of a separable polynomial contains only separable elements. In review, if G is given, the corresponding F is the fixed field under G; conversely,

if an F obtainable from a group is given, the group is recovered by taking all relative automorphisms of E/F. Moreover, F is obtainable from a group if E/F is the splitting field of a separable polynomial. Further, if in this correspondence $F_1 \longleftrightarrow G_1$ and $F_2 \longleftrightarrow G_2$, then it is easy to show that $F_1 \subset F_2 \Longleftrightarrow G_1 \supset G_2$. Thus, this one-one correspondence inverts the inclusion relation. We shall now show that $F_1 \subset F_2 \subset E$ where ~~only~~ F_1 is obtainable from a group, (i.e. F_1 is a fixed field) $\Longrightarrow F_2$ is also obtainable from a group. It suffices to show that F_2 is normal under E: since E/F_1 is the splitting field of a separable $f(x) \varepsilon F_1[x]$, it follows that E is the splitting field over F_2 of the same polynomial $f(x) \varepsilon F_1[x] \subset F_2[x]$. It is customary to state a much weaker result without the use of Γ, viz. given F and a splitting field E of a separable polynomial f(x), then any intermediate field F_1 is obtainable by a subgroup H of the group G of all relative automorphisms of E/F. This is illustrated in Fig. A where the labeled arrows indicate the corresponding group. Fig. B indicates some obvious relations when two fields F_1 and F_2 are intermediate to E and F; H is the group generated by H_1 and H_2. To continue the study of subfields, apply $\tau \varepsilon G$ to F_1 in

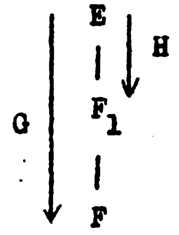


Fig. A

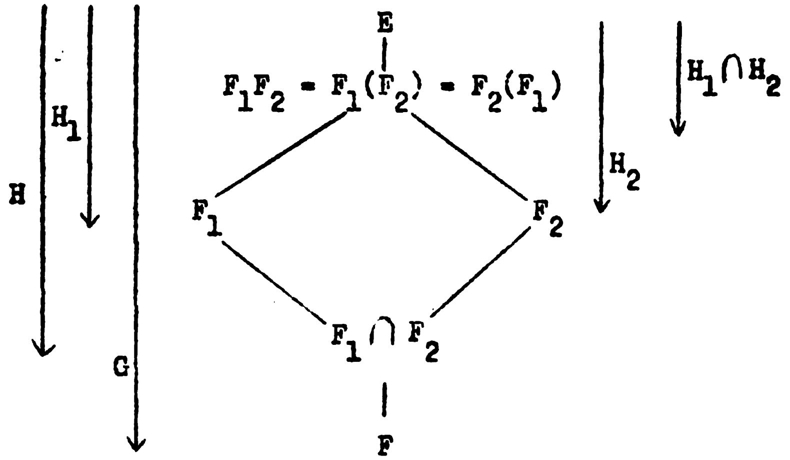


Fig. B

the situation given by Fig. A to obtain the structure indicated in Fig. C. The problem at hand is to characterize the group L. Let x_1 range over F_1, then $\lambda \epsilon L \Longleftrightarrow \lambda(\tau(x_1)) = \tau(x_1) \Longleftrightarrow \tau^{-1}\lambda\tau(x_1) = x_1 \Longleftrightarrow \tau^{-1}\lambda\tau \epsilon H \Longleftrightarrow \lambda \epsilon \tau H\tau^{-1}$. Hence, $L = \tau H\tau^{-1}$. All relative isomorphisms of F_1 can be obtained in this fashion as the relative isomorphisms of F_1 can be extended to relative automorphisms of E/F. In the case illustrated in Fig. C, $F_1 = \tau(F_1) \Longleftrightarrow H = \tau H\tau^{-1}$. When is F_1/F normal? Before the answer is given, observe in the example below that $Q(\sqrt[4]{2}, i)$ is normal over Q, but that $Q(\sqrt[4]{2})$ is not normal over Q; also observe that $Q(\sqrt{2})$ is normal over Q and $Q(\sqrt[4]{2})$ is normal over $Q(\sqrt{2})$, but $Q(\sqrt[4]{2})$ is not normal over Q. The only general statement has been proved, i.e. E normal over F $\Longrightarrow$ E normal over F_1. Contention: F_1/F normal $\Longleftrightarrow \tau(F_1) = F_1$ for all τ. Proof: a) F_1/F normal. Then F_1 is the splitting field of a certain polynomial $g(x) \epsilon F[x]$. Now, τ can only permute the roots of g(x). Hence, $\tau(F_1) = F_1$. b) Let $\tau(F_1) = F_1$ for all τ. Let p(x) be an irreducible polynomial having a root α in F_1, then all other roots are $\tau(\alpha) \epsilon F_1$. So, p(x) splits in F_1. Combining the last two results, we have F_1/F normal $\Longleftrightarrow H = \tau H\tau^{-1}$ for all $\tau \Longleftrightarrow$ H an invariant subgroup of G. Thus, if F_1 is normal over F, Fig. C collapses back to Fig. A and $\tau H\tau^{-1} = H$, or equivalently $\tau H = H\tau$, for all τ. What, now, are the different maps of F_1/F? All of these maps come from $\tau \epsilon G$; however,

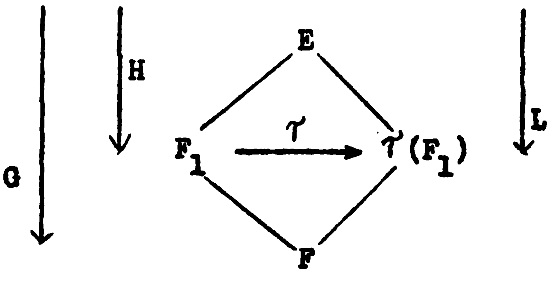


Fig. C

it may happen that τ's different in E become indistinguishable when they are restricted to F_1, e.g. the elements of H. So, when do τ_1 and τ_2 (ε G) have the same restriction to F_1? $\tau_1(x) = \tau_2(x)$ for all $x \varepsilon F_1$ $\longleftrightarrow$ $x = \tau_1^{-1}\tau_2(x)$ for all $x \varepsilon F_1$ $\longleftrightarrow$ $\tau_1^{-1}\tau_2 \varepsilon H$ $\longleftrightarrow$ $\tau_2 \varepsilon \tau_1 H$ $\longleftrightarrow$ $\tau_2 H = \tau_1 H$ $\longleftrightarrow$ τ_1 and τ_2 are in the same left coset. In particular, the number of maps = the number of cosets = $[F_1:F]$, and this holds even if F_1/F is not normal. However, to describe the Galois group of F_1/F in case F_1/F is normal, observe that its elements are restrictions of the elements of G. Each coset of H, then, provides an automorphism of F_1/F. Since the product of cosets is obtained by the multiplication of their representatives, we recognize the Galois group of F_1/F as the factor group G/H. We have seen that the following may be added to the fundamental theorem of Galois theory: A subfield (intermediate) F_1 is normal $\longleftrightarrow$ H is an invariant subgroup, and the Galois group for F_1/F will be the factor group G/H.

In this paragraph we shall study a theorem which Lagrange called the theorem of natural irrationalities. A modern interpretation of Lagrange's idea in attempting to solve an equation $f(x) = 0$ where $f(x) \varepsilon F[x]$ follows: 1) Start with F; 2) construct the splitting field E of $f(x)$; 3) replace F by a much bigger field Ω in which the problem is possibly simpler; 4) compound Ω and E; 5) characterize the Galois group G'. Certain logical difficulties are apparent. Can Ω and E be compounded? Even so, is ΩE normal over Ω? To avoid these questions, the diagram in Fig. D needs to be corrected.

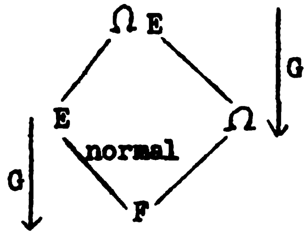


Fig. D

Consider $f(x)$ as $\varepsilon\ \Omega[x]$. Let Ω' be the splitting field of $f(x) = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n)$ over Ω. Then, let $E = F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ so that E is the splitting field of $f(x)$ over F. This E may be different from the original given splitting field but is certainly isomorphic to it by the uniqueness property of splitting fields. Now, $\Omega' = \Omega E$, and we are ready to relate G' to G. Consider the restrictions τ of the elements τ' of G' to E. τ', leaving Ω fixed, certainly leaves F fixed. Therefore, τ leaves F fixed. Since τ' permutes the roots of $f(x)$, τ maps E onto E, and τ is indeed a relative automorphism of E/F. Thus a mapping G' ⟶ G is given by $\tau' \longrightarrow \tau$. As τ' is a relative automorphism of Ω'/Ω, τ' is determined by the images of $\alpha_1, \alpha_2, \ldots, \alpha_n$. But these images are already known if τ is known, i.e. τ determines τ' and our mapping is one-one into! Moreover, the mapping is an isomorphism into as the homomorphism properties are obvious. The image H of G' is a subgroup of G. To find the fixed field corresponding to H, view H to be G' by looking only at the effect of G' in E. The fixed elements are obviously those of $\Omega \cap E$. The group H, the image of G', can therefore be described as the subgroup of G which has $\Omega \cap E$ as fixed field.

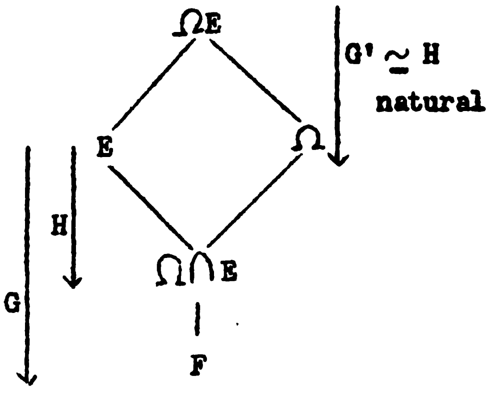


Fig. E

Example: Let F = Q, $f(x) = x^4 - 2$, and E be the splitting field of $f(x)$. $E = Q(\sqrt[4]{2}, -\sqrt[4]{2}, i\sqrt[4]{2}, -i\sqrt[4]{2}) = Q(\sqrt[4]{2}, i)$. By Eisenstein's criterion $[Q(\sqrt[4]{2}):Q] = 4$. As i satisfies a quadratic equation and as complex

numbers are introduced, $[E : Q(\sqrt[4]{2})] = 2$. Therefore, $[E:Q] = 8$ and E is normal over Q. Consequently, F is obtainable from the Galois group G of order 8. As it is sufficient in the case of automorphisms to give the images of the generators, those possible for $\sqrt[4]{2}$ are $\pm\sqrt[4]{2}$ and $\pm i\sqrt[4]{2}$ and for i, $\pm$ i. As eight must exist, these eight cases will give the Galois group. In the table that follows the elements of G are listed according to the images of $\sqrt[4]{2}$ and i; the last column gives the period of each element. Adjacent to the table, the essential relations for the arithmetic in G are given.

1	$\sqrt[4]{2}$	i	1
σ	$i\sqrt[4]{2}$	i	4
σ^2	$-\sqrt[4]{2}$	i	2
σ^3	$-i\sqrt[4]{2}$	i	4
τ	$\sqrt[4]{2}$	-i	2
$\tau\sigma$	$-i\sqrt[4]{2}$	-i	2
$\tau\sigma^2$	$-\sqrt[4]{2}$	-i	2
$\tau\sigma^3$	$i\sqrt[4]{2}$	-i	2

$$\sigma^4 = 1$$

$$\tau^2 = 1$$

$$\sigma^{-1} = \sigma^3$$

$$\tau^{-1} = \tau$$

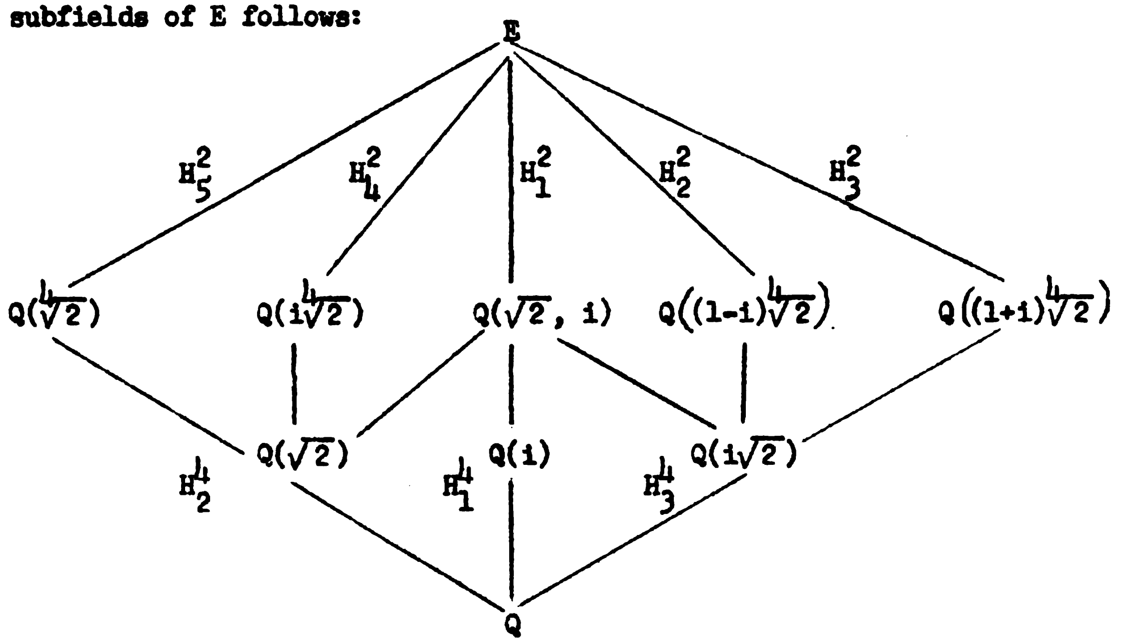
$$\sigma\tau = \tau\sigma^3 = \tau\sigma^{-1}$$

$$\sigma^2\tau = \tau\sigma^2$$

$$\sigma^3\tau = \sigma^{-1}\tau = \tau\sigma$$

Since the order of G is 8, the non-trivial subgroups must have order 2 or 4 and can be shown to be: $H_1^2 = \{1, \sigma^2\}$, $H_2^2 = \{1, \tau\sigma\}$, $H_3^2 = \{1, \tau\sigma^3\}$, $H_4^2 = \{1, \tau\sigma^2\}$, $H_5^2 = \{1, \tau\}$, $H_1^4 = \{1, \sigma, \sigma^2, \sigma^3\}$, $H_2^4 = \{1, \sigma^2, \tau, \tau\sigma^2\}$, $H_3^4 = \{1, \sigma^2, \tau\sigma, \tau\sigma^3\}$. The diagram of the

subfields of E follows:



From the subgroups the subfields may be constructed. For example, we shall show how to construct the subfield corresponding to H_2^2. For each $\theta \varepsilon E$, $\theta = a + b\sqrt[4]{2} + c(\sqrt[4]{2})^2 + d(\sqrt[4]{2})^3 + ei + fi\sqrt[4]{2} + gi(\sqrt[4]{2})^2 + hi(\sqrt[4]{2})^3$ uniquely. The desired subfield, being fixed under H_2^2, consists of all elements θ for which $\theta = \tau\sigma(\theta) = a - f\sqrt[4]{2} - c(\sqrt[4]{2})^2 + h(\sqrt[4]{2})^3 - ei - bi\sqrt[4]{2} + gi(\sqrt[4]{2})^2 + di(\sqrt[4]{2})^3$. From the uniqueness we must have b = -f, c = -c, d = h, and e = -e. Consequently, $\theta = a + b\sqrt[4]{2} + d(\sqrt[4]{2})^3 - bi\sqrt[4]{2} + gi(\sqrt[4]{2})^2 + di(\sqrt[4]{2})^3 = a + b(1 - i)\sqrt[4]{2} + gi\sqrt{2} + d(1 + i)(\sqrt[4]{2})^3$. Let $\alpha = (1 - i)\sqrt[4]{2}$.; then $\alpha^2 = -2i\sqrt{2}$ and $\alpha^3 = -2(1 + i)(\sqrt[4]{2})^3$. Hence, $\theta = a + b\alpha - \frac{g}{2}\alpha^2 - \frac{d}{2}\alpha^3 = a + b\alpha + c'\alpha^2 + d'\alpha^3$ where $\alpha^4 = -8$, and th3 subfield corresponding to H_2^2 is $Q((1 - i)\sqrt[4]{2})$.

5.4 <u>First Cohomology Group of a Field</u>. Let G be a group with elements σ, τ, ... A is a G module if

1) A is an abelian group,

2) there exists a composition $\sigma(a) \varepsilon A$ which is a homomorphism, i.e. $\sigma(a + b) = \sigma(a) + \sigma(b)$,

3) $\sigma\bigl(\tau(a)\bigr) = (\sigma\tau)a$.

For example, let A be any abelian group and $\sigma(a) = a$ for all $\sigma \varepsilon G$ (the so-called trivial action). In case the group A is multiplicative, exponential notation will be used: $\sigma(ab) = \sigma(a)\sigma(b)$ will be written as $(ab)^\sigma = a^\sigma b^\sigma$, and 3) will be written as $(a^\tau)^\sigma = a^{\sigma\tau}$. If E is a field and G a group of automorphisms of E, two less trivial examples can be formed: a) E as an additive group can be taken as A, in which case E is said to be a G module under addition; b) the multiplicative group $E - \{0\}$ can be used as A. Cohomology groups are constructed from two elements, viz. G and a G module, A. G is called the group, and the elements of A are known as the coefficients. The functions $f(\sigma_1, \sigma_2, \ldots, \sigma_p)$ on the cartesian product G^p with values in A are called p-cochains. The set of p-cochains is denoted by $C^p(G, A)$ which is abbreviated to C^p. With the compositions defined in the obvious fashion, C^p itself is an additive group. The 1-cochains are the functions $f(\sigma)$ with values in A and σ ranging over G; the 2-cochains are the functions $f(\sigma, \tau)$ with values in A and σ, τ ranging over G; etc. The elements a of A will be taken as the 0-cochains; the set of these will be called $C^0(G, A)$. For each p, the coboundary operator (mapping) $\delta : C^p \longrightarrow C^{p+1}$ is defined. If $f \varepsilon C^0$, then $\delta f \varepsilon C^1$ for which $(\delta f)(\sigma) = \sigma f - f$. If $f \varepsilon C^1$, then $(\delta f) \varepsilon C^2$ such that $(\delta f)(\sigma, \tau) = \sigma f(\tau) - f(\sigma\tau) + f(\sigma)$. If $f \varepsilon C^2$, then $(\delta f)(\sigma, \tau, \rho) = \sigma f(\tau, \rho) - f(\sigma\tau, \rho) + f(\sigma, \tau\rho) - f(\sigma, \tau)$. If $f \varepsilon C^3$, then $(\delta f)(\sigma, \tau, \rho, \lambda) =$

$\sigma f(\tau, \rho, \lambda) - f(\sigma\tau, \rho, \lambda) + f(\sigma, \tau\rho, \lambda) - f(\sigma, \tau, \rho\lambda) + f(\sigma, \tau, \rho)$. Etc. We shall verify the general result $\delta\delta f = 0$, the coboundary of the coboundary is zero, in the case important to us. For $f \epsilon C^0$, $(\delta(\delta f))(\sigma, \tau) = \sigma(\delta f)(\tau) - \delta f(\sigma\tau) + \delta f(\sigma) = \sigma(\tau f - f) - (\sigma\tau f - f) + (\sigma f - f) = 0$. Notice that $\delta(f + g) = \delta f + \delta g$. Hence, the coboundary operator $\delta : C^p \longrightarrow C^{p+1}$ is a homomorphism of the additive group C^p. The kernel of δ is the set of cochains in C^p whose coboundary is zero. These cochains are called cocycles. Z^p will be used to designate the kernel of δ in C^p. To determine the cocycles in Z^0 let $f \epsilon A$ where $\delta f = 0$. This means $(\delta f)(\sigma) = \sigma f - f = 0$ and consequently $\sigma f = f$ for all σ. So usually the elements of A are not cocycles but just cochains. For instance, if G is the Galois group of a field and the module is the field, then the cocycles are the elements of the fixed field. Important for our purposes will be the following: $f \epsilon Z^1 \longleftrightarrow (\delta f)(\sigma, \tau) = 0 \longleftrightarrow \sigma f(\tau) - f(\sigma\tau) + f(\sigma) = 0$. In terms of the structure indicated in Fig. A, the p-dimensional cohomology group $H^p(G, A)$ is defined as the factor group Z^p/B^p where B^p is the image of C^{p-1} under δ. The current interest in algebra is primarily in dimensions from 0 to 3.

$$\begin{tikzcd}
C^{p-1} \arrow[r] & C^p \arrow[r] \arrow[d, dash] & C^{p+1} \\
 & Z^p \arrow[d, dash] & \\
 & B^p &
\end{tikzcd}$$

Fig. A

We shall now show that $H^1(G, A) = 0$ for G the Galois group of the field E and A = E (additive) or A = E − {0} (multiplicative). The additive case will be discussed first. Contention: $H^1(G, E) = 0$, i.e. $Z^1/B^1 = 0$ or, equivalently, $Z^1 = B^1$. Since $B^1 \subset Z^1$, we have only to show $Z^1 \subset B^1$. It will suffice to show that

(*) $\sigma f(\tau) - f(\sigma\tau) + f(\sigma) = 0 \Longrightarrow$ the existence of a ε E such that $f(\sigma) = \sigma a - a$. Multiplication of (*) by $\sigma\tau(\theta)$ where θ ε E gives $\sigma\tau(\theta)\cdot\sigma f(\tau) - \sigma\tau(\theta)\circ f(\sigma\tau) + \sigma\tau(\theta)\cdot f(\sigma) = 0$. Since $\sigma\tau(\theta)\cdot\sigma f(\tau)$ can be replaced by $\sigma\big(\tau(\theta)\cdot f(\tau)\big)$, summing this last equation over τ gives

$$\sigma\sum_{\tau}\tau(\theta)\,f(\tau) - \sum_{\tau}\underbrace{\sigma\tau}_{\tau}(\theta)\,f(\underbrace{\sigma\tau}_{\tau}) + f(\sigma)\sum_{\tau}\underbrace{\sigma\tau}_{\tau}(\theta) = 0$$ where

replacements permissible by the group property are indicated below the braces. Letting $\alpha = \sum_{\tau}\tau(\theta)\,f(\tau)$ and noting that $\sum_{\tau}\tau(\theta) = S(\theta)$ (trace), we have $\sigma(\alpha) - \alpha + f(\sigma)\cdot S(\theta) = 0$. By choosing θ such that $S(\theta) \neq 0$, it follows that $f(\sigma) = \dfrac{\alpha - \sigma(\alpha)}{S(\theta)} = \sigma\left(-\dfrac{\alpha}{S(\theta)}\right) - \left(-\dfrac{\alpha}{S(\theta)}\right)$ since the trace of any element belongs to the fixed field. Hence, $-\dfrac{\alpha}{S(\theta)}$ is the desired a ε E. We come now to the multiplicative case. Take A = E − {0} (multiplicative) and G as a group of automorphisms of the field E. For $f \in C^1$, we now have $\delta f(\sigma,\tau) = f(\tau)^{\sigma}\,f(\sigma\tau)^{-1}f(\sigma)$, and, in particular, f a cocycle ($\in Z^1$) means $\delta f(\sigma,\tau) = 1$, or equivalently, $f(\sigma\tau) = f(\tau)^{\sigma}\,f(\sigma)$. Upon multiplication by $\theta^{\sigma\tau}$ where θ ε E, this becomes $\theta^{\sigma\tau}\,f(\sigma\tau) = \left(\theta^{\tau}f(\tau)\right)^{\sigma}\cdot f(\sigma)$. Addition over τ gives

$$\sum_{\tau}\theta^{\overbrace{\sigma\tau}^{\tau}}f(\overbrace{\sigma\tau}^{\tau}) = f(\sigma)\left(\sum_{\tau}\theta^{\tau}f(\tau)\right)^{\sigma}$$

where the image of the sum under the automorphism σ replaces the sum of the images and replacements permissible by the group property are indicated above the braces. We can solve for $f(\sigma)$ and put

$a^{-1} = \sum\limits_{\tau} \theta^{\tau} f(\tau)$, if we have for some $\theta \varepsilon E$ that $\sum \theta^{\tau} f(\tau) \neq 0$. Then

$$f(\sigma) = \frac{\sum\limits_{\tau} \theta^{\tau} f(\tau)}{\left(\sum\limits_{\tau} \theta^{\tau} f(\tau)\right)^{\sigma}} = \frac{a^{\sigma}}{a} .$$

Otherwise $\sum\limits_{\tau} f(\tau)\, \theta^{\tau} = 0$ for all $\theta \varepsilon E$, but this is not possible as all the automorphisms are linearly independent. Hence, $Z^1 \subset B^1$, which suffices as in the additive case. Let us see what happens in the special case that E is a quadratic extension field, i.e. the Galois group G is of order 2. Let $G = \{1, \sigma\}$ where $\sigma^2 = 1$. Evidently, $f \varepsilon C^1$ is characterized by the two values $f(1)$ and $f(\sigma)$. To see which of these 1-cochains are cocycles, let σ and τ in the condition $f(\sigma\tau) = f(\tau)^{\sigma} f(\sigma)$ take on all values possible: $f(1)^1 f(1) = f(1)$, $f(\sigma) = f(\sigma)^1 f(1)$, $f(\sigma) = f(1)^{\sigma} f(\sigma)$, $f(1) = f(\sigma)^{\sigma} f(\sigma)$. This means, for f a cocycle, $f(1) = 1$ and $f(\sigma) = \alpha \varepsilon E$ such that $\alpha\alpha^{\sigma} = 1$. The product obtained by applying all of the automorphisms to α and multiplying, $\alpha\alpha^{\sigma}$ in this example, is called the norm of α, $N(\alpha)$. There exists $a \varepsilon E - \{0\}$ such that $f(\tau) = \frac{a^{\tau}}{a}$ for every $\tau \varepsilon G$. For $\tau = 1$, this gives only the trivial statement $1 = 1$; however, for $\tau = \sigma$, this requires that $\alpha = \frac{a^{\sigma}}{a}$. Independently of the notion of cocycles, we can say that an element α whose norm is 1 ($\alpha\alpha^{\sigma} = 1$) must have the form $\alpha = \frac{a^{\sigma}}{a}$. If E is the complex number field and G consists of the two obvious automorphisms, i.e. $\sigma(x + yi) = x - yi$, then our numbers α where $\alpha = \frac{x - yi}{x + yi}$ are the numbers

on the unit circle. Unless $\alpha = -1$, a^{-1} can be taken as $a^{-1} = 1 + \alpha$ by putting $\theta = 1$. This method is even more useful for other quadratic extension fields.

Consider the more general case where G is any cyclic group of order n with generator σ and A is any G module, which we shall write additively. The cocycle equation is $f(\sigma^{i+j}) = \sigma^{i}f(\sigma^{j}) + f(\sigma^{i})$. For $i = j = 0$, this gives $f(1) = f(1) + f(1)$ and, therefore, $f(1) = 0$. Call $f(\sigma) = \alpha$. For $i = j = 1$, $f(\sigma^2) = \alpha + \sigma\alpha$; for $i = 2$ and $j = 1$, $f(\sigma^3) = \alpha + \sigma\alpha + \sigma^2\alpha$; for $j = 1$ and arbitrary i, $f(\sigma^k) = \alpha + \sigma\alpha + \sigma^2\alpha + \ldots + \sigma^{k-1}\alpha$ where $k = i + j$. For $k = n$, this means $0 = \alpha + \sigma\alpha + \sigma^2\alpha + \ldots + \sigma^{n-1}\alpha$ or, equivalently, $0 = S(\alpha)$ as $\sigma^n = 1$ and $f(1) = 0$. In other words, for the cyclic group G, a cocycle is determined by α, which must be selected such that its trace is 0. Conversely, does any number whose trace is 0 lead to the cocycle equation? Select α with $S(\alpha) = 0$ $\big(N(\alpha) = 1$ in case A is multiplicative$\big)$. Define $f(\sigma^k) = \alpha + \sigma\alpha + \sigma^2\alpha + \ldots + \sigma^{k-1}\alpha$. That this definition is well-defined, i.e. that $f(\sigma^k) = f(\sigma^{k+n})$, follows from: $f(\sigma^{k+n}) - f(\sigma^k) = \sigma^k\alpha + \sigma^{k+1}\alpha + \ldots + \sigma^{k+n-1}\alpha = S(\alpha) = 0$. It is a triviality to show that f so-defined is a cocycle. If we make the further assumption that $H^1(G, A) = 0$, then there exists an $a \varepsilon A$ such that $f(\sigma^k) = \sigma^k a - a$ for all k. For $k = 0$, this is trivial; but for $k = 1$, we have $\alpha = f(\sigma) = \sigma a - a$; and, in general, $\alpha + \sigma\alpha + \ldots + \sigma^{k-1}\alpha = \sigma^k a - a$. However, it is easy to show that the general result is a consequence of $\alpha = \sigma a - a$. So in case the group G is cyclic the first cohomology group deals with elements ($\varepsilon\ Z^1$) of trace zero. Should the

first cohomology group be trivial, each element α with trace 0 has the form $\sigma a - a$. Conversely, $\alpha = \sigma a - a$ obviously has trace zero as $S(\alpha) = S(\sigma a) - S(a)$ and $S(\sigma a) = S(a)$. If the G module is written multiplicatively, this would mean an element with norm 1 would have the form $\frac{a^\sigma}{a}$. Let us return now to Galois theory where we have proved the first cohomology group to be trivial in both the additive and multiplicative cases. If the field extension is cyclic (i.e. if the Galois group G is cyclic), it follows that the elements having trace 0 or norm 1 are exactly those obtainable by $\sigma a - a$ or $\frac{a^\sigma}{a}$ respectively. The result for norm 1 is referred to as Hilbert Theorem 90.

Application: Suppose E is a cyclic extension of the field F, $[E:F] = n$, and σ is a generator of the Galois group. Assume further that the ground field F contains all of the primitive n-th roots of unity, i.e. $x^n - 1$ splits into distinct factors. This would not be the case if the characteristic of the field were to divide n. Hence, the characteristic p is either 0 or a prime which does not divide n. Let ζ be a primitive n-th root of unity, i.e. ζ not an m-th root of unity for $m < n$. As $\zeta \varepsilon F$ (fixed), all images of ζ are in F, and $N(\zeta) = \zeta^n = 1$. By Hilbert Theorem 90, $\zeta = \frac{\sigma a}{a}$ for some $a \varepsilon E$. This means that the σ image of a is a multiplied by the root of unity: $\sigma a = \zeta \cdot a$. Also, $\sigma^2 a = \zeta^2 \cdot a$, $\sigma^3 a = \zeta^3 \cdot a$, etc. So the images of a under the automorphisms are: $a, \zeta a, \zeta^2 a, \ldots, \zeta^{n-1} a$. These images are distinct as ζ is primitive. Consequently, the irreducible equation for a over F is of degree n. It follows that $E = F(a)$. Moreover, $\sigma(a^n) = a^n$, i.e. a^n is fixed under the generator of G and, therefore, under G. Thus, $a^n = b \varepsilon F$ and a is a root of $x^n - b = 0$. We can conclude that the field E is obtainable from F by a radical, $E = F(\sqrt[n]{b})$.

For the converse of the last result, let F be a ground field containing the primitive n-th roots of unity. Consider $E = F(\sqrt[n]{a})$ where $a \varepsilon F$. The characteristic $p \nmid n$; otherwise there would not be any primitive n-th roots of unity. E splits $x^n - a$ and is separable because the n-th roots of unity in this case are distinct. Hence E is normal. If σ belongs to the Galois group G, then σ sends $\sqrt[n]{a}$ into another root of the same equation, i.e. $\sigma(\sqrt[n]{a}) = \zeta_\sigma \sqrt[n]{a}$ where $\zeta_\sigma^n = 1$. Consequently, we may consider the mapping of G into the multiplicative group of the n-th roots of unity given by $\sigma \longrightarrow \zeta_\sigma$. As ζ_σ determines σ, this mapping is one-one into. Furthermore, since $\zeta_{\sigma\tau}(\sqrt[n]{a}) = \sigma\tau(\sqrt[n]{a}) = \sigma(\tau\sqrt[n]{a}) = \sigma(\zeta_\tau\sqrt[n]{a}) = \zeta_\tau\zeta_\sigma\sqrt[n]{a}$ (ζ_τ in ground field), $\zeta_{\sigma\tau} = \zeta_\sigma\zeta_\tau$. Therefore the map is a homomorphism and, consequently, an isomorphism into. It follows that G is isomorphic to a subgroup of the multiplicative group of the n-th roots of unity and, hence, is a subgroup of a cyclic group of order n. Suppose $1, \zeta, \ldots, \zeta^{n-1}$ are the n-th roots of unity, and consider the set S of all integers r such that ζ^r is in the subgroup isomorphic to G. S is an ideal of Z for $\zeta^r, \zeta^s \varepsilon$ subgroup $\longrightarrow \zeta^{r \pm s} \varepsilon$ subgroup. As every ideal in Z is principal, we can let $S = dZ$. Since $\zeta^n \varepsilon$ subgroup, $n \varepsilon dZ$ and $d|n$. Therefore, the subgroup consists of $1, \zeta^d, \zeta^{2d}, \ldots, \zeta^{(e-1)d}$ where $n = de$. Thus, if F is a ground field containing the primitive n-th roots of unity and if $E = F(\sqrt[n]{a})$ where $a \varepsilon F$, then the Galois group is cyclic and has as order a divisor of n.

We shall consider now the case $n = p$. Let F be a ground field with prime characteristic p. Let E/F be normal with $[E:F] = p$. As the order of the Galois group G is a prime, the group must be cyclic. Take σ as a generator of G. Since $x^p - 1 = (x - 1)^p$, using the primitive roots of

unity is hopeless. However, a similar method will be used. For the purpose of stating more elegant theorems later, it is useful to consider this method also as "solvable by radicals." To see which equations constitute the modified radical type, we turn again to cocycles. Here we use the additive notation and choose the element 1, which has trace zero: $S(1) = 1 + \sigma(1) + \sigma^2(1) + \ldots + \sigma^{p-1}(1) = 1 + 1 + \ldots + 1 = 0$. Consequently, there exists an $\alpha \varepsilon E$ such that $1 = \sigma(\alpha) - \alpha$, and we have the following p elements, which are distinct: $1(\alpha) = \alpha + 0$, $\sigma(\alpha) = \alpha + 1$, $\sigma^2(\alpha) = \alpha + 2, \ldots, \sigma^{p-1}(\alpha) = \alpha + (p - 1)$. α is moved by the group into these p distinct elements. It follows that $Irr(\alpha, F)$ has p roots and $[F(\alpha):F] = p$. Therefore, $E = F(\alpha)$. To find what equation α satisfies, call $a = \alpha^p - \alpha$ and compute $\sigma(a)$ using $\sigma(\alpha) = \alpha + 1$: $\sigma(a) = (\alpha + 1)^p - (\alpha + 1) = \alpha^p + 1 - \alpha - 1 = \alpha^p - \alpha = a$. Hence σ and, consequently, G leave a fixed. In other words, $a \varepsilon F$ and α is a root of the equation $x^p - x - a = 0$. Any solution of this equation is called a <u>modified radical</u>. Observe that: α solves $\alpha^p - \alpha - a = 0$ and β solves $\beta^p - \beta - b = 0 \Longrightarrow \alpha + \beta$ solves $(\alpha + \beta)^p - (\alpha + \beta) - (a + b) = 0$, i.e. the modified p-th root is a linear operator. This is even simpler than the analogous property in the other case, vis. the p-th root of a product is the product of the p-th roots. Conversely, given a field F with prime characteristic p and $f(x) = x^p - x - a \varepsilon F[x]$, does the solution give a cyclic field? We shall make use of the periodic nature of f(x), i.e. $f(x + 1) = (x + 1)^p - (x + 1) - a = x^p - x - a$. Let $F(\alpha)$ be such that $\alpha^p - \alpha - a = 0$. Then $f(\alpha) = 0 \Longrightarrow f(\alpha + 1) = 0 \Longrightarrow f(\alpha + 2) = 0 \Longrightarrow \ldots \Longrightarrow f(a + p - 1) = 0$. Hence, $\alpha, \alpha + 1, \alpha + 2, \ldots, \alpha + p - 1$

are p distinct roots of $f(x)$, and $f(x) = \prod_{i=0}^{p-1}\bigl(x - (\alpha + i)\bigr)$.

Consequently, $f(x)$ is separable. $F(\alpha)$ is the splitting field of $f(x)$ and, therefore, is normal. To decide whether or not $f(x)$ is irreducible over F, suppose $g(x)$ is irreducible and $g(x)|f(x)$. Replacing x by $x + 1$ gives $g(x+1)|f(x+1) = f(x)$. We shall distinguish between two cases:

1) $g(x) = g(x + 1)$. This can happen only if $g(x) = f(x)$ since α a root of $g(x) \Longrightarrow g(x)$ has the p distinct roots α, $\alpha + 1$, $\alpha + 2$, ..., $\alpha + p - 1$ and, therefore, has degree p.

2) $g(x) \neq g(x + 1)$. In this case, $g(x)$, $g(x + 1)$, ..., $g(x + p - 1)$ are distinct and $g(x)g(x + 1) \dots g(x + p - 1)|f(x)$. Therefore, $g(x)$ must be linear.

It follows that $[F(\alpha):F] = 1$ or p. We have proved the following theorem: If F is a ground field of prime characteristic p, then E/F is normal of degree p (and the Galois group is cyclic of order p) $\Longleftrightarrow$ $E = F(\alpha)$ where α is a root of $f(x) = x^p - x - a \in F[x]$ and $f(x)$ is irreducible over F.

Let F be a ground field whose characteristic $p \nmid n$, and let E be the splitting field of $f(x) = x^n - 1$. If ζ is a root of $f(x)$, then ζ is not a double root, since $x^n - 1 = x^n - \zeta^n = (x - \zeta)(x^{n-1} + \zeta x^{n-2} + \dots + \zeta^{n-1})$ and for $x = \zeta$ $\quad x^{n-1} + \zeta x^{n-2} + \dots + \zeta^{n-1} = n\zeta^{n-1} \neq 0$. Therefore, E/F is normal. The roots of $f(x)$ form a cyclic multiplicative group of order n which can be represented by $1, \zeta, \zeta^2, \dots, \zeta^{n-1}$. Moreover, $E = F(\zeta)$. The problem is: What can be said about the Galois group G? If $\sigma \in G$, then $\sigma(\zeta)$ is ~~another~~ primitive n-th root of unity, as an automorphism preserves the essential features of multiplication. Therefore, $\sigma(\zeta) = \zeta^{m_\sigma}$ where m_σ is only defined mod n (actually a residue class) and is relatively prime to n. The mapping of G into the residue class group Z modulo nZ given by $\sigma \longrightarrow m_\sigma + nZ$ is one-one.

Furthermore, it is a homomorphism for $\zeta^{m_{\sigma\tau}} = \sigma\tau(\zeta) = \sigma(\zeta^{m_\tau}) =$ $(\sigma(\zeta))^{m_\tau} = (\zeta^{m_\sigma})^{m_\tau} = \zeta^{m_\sigma m_\tau}$ and $m_{\sigma\tau} + nZ = (m_\sigma + nZ)(m_\tau + nZ)$. Therefore, the map is an isomorphism of G onto a subgroup of the prime residue classes mod n. It will be especially important for our purposes that G is commutative.

5.5 Towers of Fields. A finite sequence of fields $F = E_0 \subset E_1 \subset E_2 \subset \dots \subset E_r = E$ is called a tower of fields. If $E_i = E_{i-1}(\alpha_i)$ such that α_i is obtained from E_{i-1} by extracting a root, the tower is called a radical tower. It suffices to consider only cases where α_i is a root of $x^{p_i} - a_i$, p_i a prime, $a_i \in E_{i-1}$, for characteristic $\neq p_i$ and α_i is a root of $x^{p_i} - x - a_i$, $a_i \in E_{i-1}$, for characteristic $= p_i > 0$. A tower in which E_i/E_{i-1} is cyclic of prime degree for all i is known as a cyclic tower. In a cyclic tower E_i/E_{i-1} is especially normal. If E/F is normal and the group G is abelian, then E/F is a cyclic tower. Proof: Every subgroup of G (abelian) is invariant. Moreover, each factor group is also abelian as the multiplication of the equivalence classes is defined in terms of representative elements. Select an element $\sigma \neq 1$ in G. Call its period d so that σ generates the subgroup $\{1, \sigma, \sigma^2, \dots, \sigma^{d-1}\}$. Let $p|d$ where p is a prime, and put $\tau = \sigma^{d/p}$ so that τ has period p. Put $H = \{1, \tau, \tau^2, \dots, \tau^{p-1}\}$, which is a cyclic subgroup of order p, and let the subfield fixed under H be called E', $F \subset E' \subset E$. Since H is invariant, E'/E is normal, and its Galois group is G/H. The proof may be completed by continuing in this manner or by using induction.

In a certain weak sense, which we shall explain, the notion of a cyclic tower is equivalent to that of a radical tower. Contention: Any radical tower is a subfield of a cyclic tower, and any cyclic tower is a subfield of a radical tower. We shall first show that every cyclic tower is part of a radical tower. Suppose we have the cyclic tower $F = E_0 \subset E_1 \subset \dots \subset E_r$ where $[E_i : E_{i-1}]$ is the prime number p_i. To obtain a radical tower, adjoin to E_r all p_i-th roots of unity for which $p_i \neq$ the characteristic, and let $K = E_r(\zeta_1, \zeta_2, \dots, \zeta_s)$ where the ζ_j are the roots of unity adjoined. Clearly, $E_r \subset K$. To show that K is a radical tower, let $K_1 = F(\zeta_1)$, $K_2 = K_1(\zeta_2)$, $K_3 = K_2(\zeta_3)$, ..., $K_s = K_{s-1}(\zeta_s)$, $K_{s+1} = K_sE_1 \big(= K_s(E_1) = E_1(K_s)\big)$, $K_{s+2} = K_{s+1}E_2$, ..., $K_{s+r} = K_{s+r-1}E_r = K$. $K_1 \subset K_2 \subset \dots \subset K_s$ is obviously a radical tower for the ζ_i are radicals of the simplest type. Consider $K_{s+1}/K_s = K_sE_1/K_s$ (see Fig. A). By the Lagrange theorem of natural irrationalities, H_1 is a subgroup of the cyclic group G_1 of order p_1. Therefore, either H_1 is the identity or H_1 is cyclic of order p_1. In the former case, $E_1K_s = K_s$ and the step is superfluous. In the latter case, since the ground field K_s contains the p_1-th roots of unity, K_{s+1} is obtainable by a radical. A continuation of this process shows that K is a radical tower. For the other part of the contention, viz. each radical tower is contained in a cyclic tower, suppose $F = E_0 \subset E_1 \subset E_2 \subset \dots \subset E_{r-1} \subset E_r$ is a radical tower. As

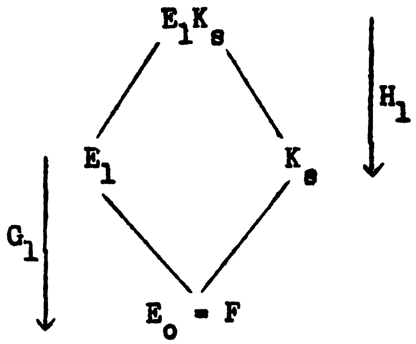


Fig. A

before, put $K = E_r(\zeta_1, \zeta_2, \ldots, \zeta_s)$, $K_1 = F(\zeta_1)$, $K_2 = K_1(\zeta_2)$, ..., $K_s = K_{s-1}(\zeta_s)$. Since K_i/K_{i-1} is obtained by adjoining a root of unity, it is abelian and, therefore, a cyclic tower (intermediate fields can be inserted if necessary to make each extension cyclic of prime degree). Thus, K_s/F is a cyclic tower. Let $K_{s+i} = K_{s+i-1}E_i$ for $i = 1, 2, \ldots, r$. For instance, $K_{s+1} = K_sE_1$ where $E_1 = E_0(\alpha_1)$ and α_1 is a root of $x^{p_1} - a_1 = 0$ or $x^{p_1} - x - a_1 = 0$ as the case may be. In either event, the group of K_{s+1}/K_s is cyclic as $K_{s+1} = K_s(\alpha_1)$. As a similar discussion holds for each K_{s+i}/K_{s+i-1}, $K_{s+r} = K$ is a cyclic tower over F, and obviously $K \supset E_r$. In line with our vague notion of an equation solvable by radicals, we agree to the following precise definition: An equation $f(x) = 0$ where $f(x)$, irreducible, ε $F[x]$ is solvable by radicals provided there exists a radical tower such that one of the elements in the tower solves the equation. It is regarding this definition that the notions of cyclic tower and radical tower are equivalent, for it follows that an element of a radical tower solves the equation if, and only if, an element of a cyclic tower does likewise. A difficulty arises in case the radical tower solving f(x) is not a normal field. Consequently, for application of Galois theory, we need the following theorem, whose proof will be temporarily postponed: Every cyclic tower is contained in another cyclic tower that is normal. By virtue of this theorem, f(x) is solvable by radicals $\Longleftrightarrow$ there exists a normal cyclic tower in which f(x) has a root. Being normal, this cyclic tower will even contain all roots of f(x). In

Fig. B let $E = E_r$ be the normal cyclic tower which solves $f(x) = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n)$, and let the corresponding groups be as indicated. Considering only E and F, we can insert the intermediate field $E' = F(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

$$\begin{tikzcd}
{} \arrow[dddd, "G"'] & {} \arrow[ddd, "G_1"] & E_r = E \arrow[d, dash] & {} \arrow[d, "G_{r-1}"] & \downarrow G_r = \{1\} \\
{} & {} & E_{r-1} \arrow[d, dash] & {} & {} \\
{} & {} & \vdots \arrow[d, dash] & {} & {} \\
{} & {} & E_1 \arrow[d, dash] & {} & {} \\
{} & {} & E_o = F & {} & {}
\end{tikzcd}$$

Fig. B

$$\begin{tikzcd}
{} \arrow[d, "H"'] & E \arrow[d, dash] & \downarrow \{1\} & {} \arrow[dd, "G"] \\
{} & E' = F(\alpha_1, \alpha_2, \ldots, \alpha_n) \arrow[d, dash] & {} & {} \\
{} & F & {} & {}
\end{tikzcd}$$

Fig. C

Fig. C indicates this as well as the corresponding groups.

In Fig. D, E_{i+1}/E_i is cyclic and especially normal. Therefore, G_{i+1} is an invariant subgroup of G_i. Moreover, G_i/G_{i+1} is cyclic and of prime degree. Hence, the groups in Fig. B form a chain which cannot be refined. Moreover, as E'/F is normal, H is an invariant subgroup and the Galois group of E'/F is G/H. The groups in Fig. C form a second chain of

$$\begin{tikzcd}
{} \arrow[dd, "G_i"'] & E \arrow[d, dash] & {} \arrow[d, "G_{i+1}"] \\
{} & E_{i+1} \arrow[d, dash] & {} \\
{} & E_i & {}
\end{tikzcd}$$

Fig. D

invariant subgroups, which can be refined to get precisely the same factor groups as in the first chain. It follows that $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ is a cyclic tower over F. Thus, if $f(x)$ is solvable by radicals, the splitting field $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ must be a cyclic tower and its composition series will contain only cyclic groups of prime order. The converse is obvious.

We shall now prove: Every cyclic tower is contained in another cyclic tower that is normal. Let E/F be separable, $E = F(\alpha_1, \alpha_2, \dots, \alpha_r)$ where the α_i are separable. We shall describe the smallest normal field containing E. Take $p_i = \mathrm{Irr}(\alpha_i, F)$. Put $f(x) = \prod_i p_i(x)$ so that each α_i is a root of $f(x)$ (there may be other roots). Consider $f(x) \varepsilon F[x] \subset E[x]$, and let K/E be the splitting field of $f(x)$. Then K/F is also the splitting field of $f(x)$ as K is obtained by adjoining all roots of $f(x)$, especially all α_i. Thus, F (all roots) $= E$ (all roots) $= K$. Let G be the group of K/F, and apply all automorphisms of G to E, i.e. for each $\sigma \varepsilon G$, consider $\sigma(E)$. $\sigma(E)$ is isomorphic to E and is contained in K. Combine all $\sigma(E)$ by defining the compositum $K_0 = \prod_\sigma \sigma(E)$. Contention: $K_0 = K$. That $K_0 \subset K$ is trivial. To show $K_0 \supset K$, let β_i be any root of p_i. Then, there exists a σ such that $\beta_i = \sigma(\alpha_i)$, and $\beta_i \varepsilon \sigma(E)$ since $\alpha_i \varepsilon E$. This means that β_i is an element of K_0, which consequently splits $f(x)$. Hence, $K_0 \supset K$. If E is a cyclic tower, each $\sigma(E)$ is also a cyclic tower, since σ maps E isomorphically onto $\sigma(E)$. To complete the proof we need that K is a cyclic tower, i.e. that the compositum of several fields each of which is a cyclic tower is also a

$$\begin{tikzcd}
& K & \\
E \arrow[ur, dash] \arrow[rr] & & \sigma(E) \arrow[ul, dash] \\
& F \arrow[ul, dash] \arrow[ur, dash] &
\end{tikzcd}$$

Fig. E

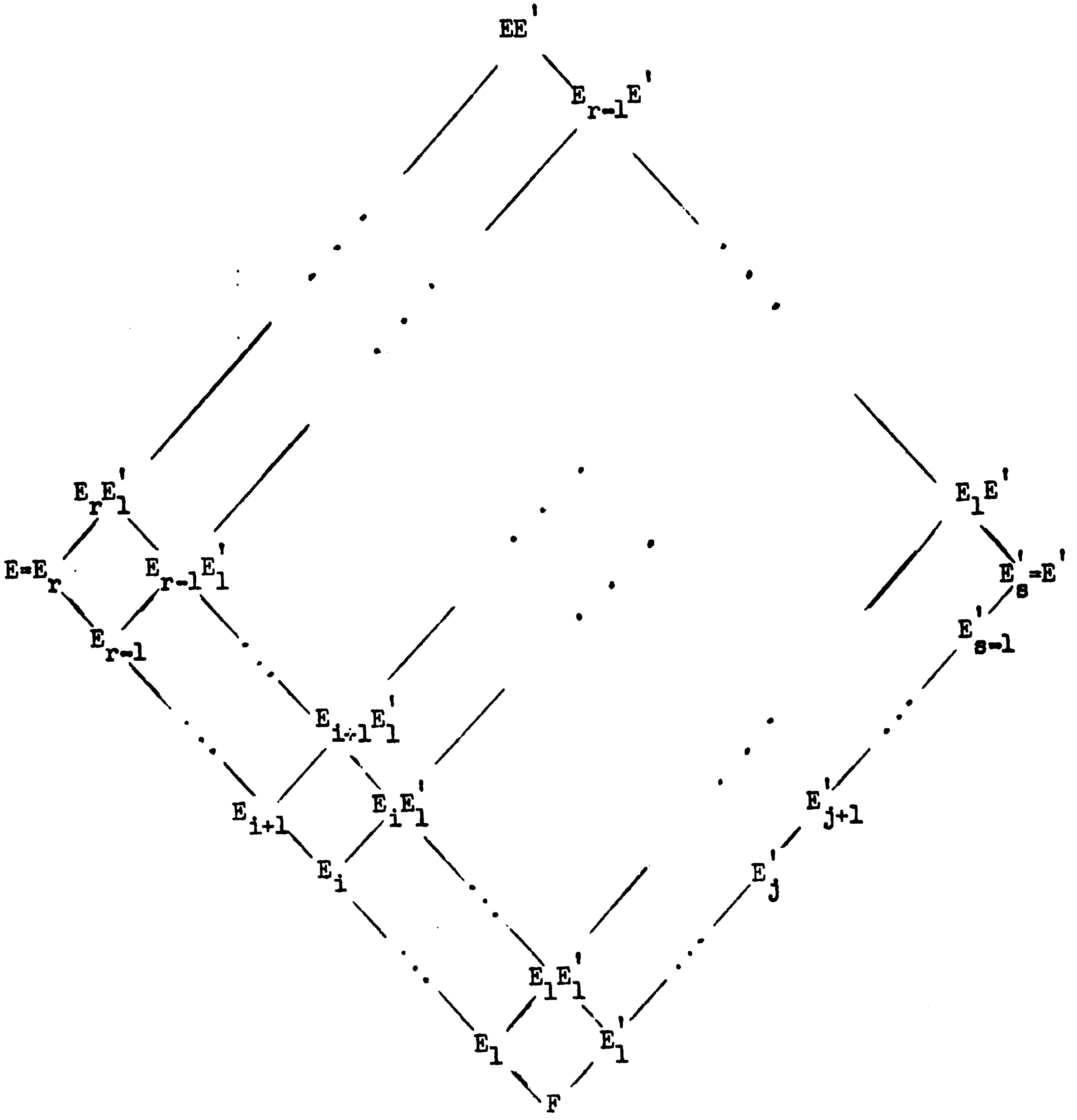

EE'
E_{r-1}E'
E_rE_1'
E=E_r
E_{r-1}E_1'
E_{r-1}
E_1E'
E_s'=E'
E'_{s-1}
E_{i+1}E_1'
E_{i+1}
E_iE_1'
E'_{j+1}
E_i
E'_j
E_1E_1'
E_1
E_1'
F

Fig. F

cyclic tower. It will suffice to show this for the compositum of two cyclic towers. Let E/F and E'/F be cyclic towers contained in the same field, then the contention is that EE'/F is also a cyclic tower. Consider the diagram in Fig. F where the generic pattern is indicated in Fig. G. We know that the groups in the tower E/F are cyclic of prime degree; however, the result can be made more general by requiring only that they be simple. As E'/F is a cyclic tower, our result will follow from showing EE'/E' is a cyclic tower. Indeed, it will follow that the groups in the tower EE'/E' form a subset of the groups in the tower E/F after an application of the Lagrange theorem of natural irrationalities to the generic pattern. This application will simply give that $G'_{i+1} = G_{i+1}$ or 1. To see this, we need to put in (see Fig. H) the intersection $E_{i+1}E'_j \cap E_iE'_{j+1}$, which is normal over $E_iE'_j$ (it is easy to show that the intersection of two normal fields is normal). Therefore, H is invariant in G_{i+1}. As G_{i+1} is simple, H = 1 or G_{i+1}. But, by Lagrange's theorem, $H \simeq G'_{i+1}$ and the proof is completed. The proof was made more general to indicate that the problem of solvability by radicals can be amended to solvability by certain other

$$\begin{tikzcd}
 & & E_{i+1}E'_{j+1} \arrow[dl, dash] \arrow[dr, dash] & & {} \arrow[d, "G'_{i+1}"] \\
{} \arrow[d, "G_{i+1}"'] & E_{i+1}E'_j \arrow[dr, dash] & & E_iE'_{j+1} \arrow[dl, dash] & {} \\
{} & & E_iE'_j & &
\end{tikzcd}$$

Fig. G

$$\begin{tikzcd}
 & & E_{i+1}E'_{j+1} \arrow[dl, dash] \arrow[dr, dash] & \\
{} \arrow[dd, "H"'] & E_{i+1}E'_j \arrow[dr, dash] & & E_iE'_{j+1} \arrow[dl, dash] \\
 & & E_{i+1}E'_j \cap E_iE'_{j+1} \arrow[d, dash] & \\
{} & & E_iE'_j &
\end{tikzcd}$$

Fig. H

specified tools, e.g, with a given set of simple groups. Returning to our specific problem, we have seen that solvability by radicals $\Longleftrightarrow$ splitting field is a cyclic tower $\Longleftrightarrow$ the composition series is cyclic of prime order. The groups that satisfy this condition are called *solvable groups*.

Cyclic towers occur in other cases besides solvability, e.g, in the problem of constructibility the splitting field should be a quadratic tower. In particular, the construction of the p-gons (p a prime) is equivalent to the construction of $\cos \frac{2\pi}{p}$. As $\zeta + \zeta^{-1} = 2 \cos \frac{2\pi}{p}$ (quadratic in ζ) where ζ is the root of unity $\cos \frac{2\pi}{p} + i \sin \frac{2\pi}{p}$, the constructibility of the p-gon means that ζ is also reached by quadratic steps. It follows from example B, p. 36, that $[Q(\zeta):Q] = p - 1$. Since we have seen that $Q(\zeta)/Q$ is abelian and is a cyclic tower, $p - 1$ must be a power of 2. Conversely, if $p - 1$ is a power of 2, the field, being abelian, is a quadratic tower and can be constructed. Therefore, a p-gon is constructible $\Longleftrightarrow$ $p = 2^i + 1$; it is also necessary that i have the form $i = 2^k$, for $i = tj$ where t is odd $\Longrightarrow (2^j + 1) | (2^{tj} + 1)$ (Gauss). The numbers $2^{2^k} + 1$ are called Fermat numbers; Fermat, however, conjectured falsely that they are all primes. For k = 0, 1, 2, 3, 4, the Fermat numbers are, respectively, the primes 3, 5, 17, 257, 65537. Euler showed that $641 | 2^{2^5} + 1$. Primes of the form $2^{2^k} + 1$ are called Fermat primes. The only n-gons constructible are those for which $n = 2^r \cdot p_1 p_2 \dots p_s$ where the p_j are different Fermat primes.

5.6 The General Equation of Degree n. Let k be any field, and consider as ground field $F = k(a_1, a_2, \dots, a_n)$ where the a_i are independent variables. If $f(x) = x^n + a_1x^{n-1} + \dots + a_n$, then $f(x) = 0$ is called the general equation of degree n. We shall prove Abel's result: The general equation of degree n is not solvable by radicals for $n > 4$. It should be remarked that every special equation may be solvable by radicals even though the general equation is proved not solvable, e.g. for k the complex number field. Let E be the splitting field of f(x) over F with $f(x) = (x - \xi_1)(x - \xi_2) \dots (x - \xi_n)$. $E = F(\xi_1, \xi_2, \dots, \xi_n) = k(\xi_1, \xi_2, \dots, \xi_n)$, the last equality holding since each a_i is a symmetric function of the ξ_j. In the example on p. 70, we studied the field $k(x_1, x_2, \dots, x_n)$ where the x_i are independent variables. For the present purposes, the following notation will be adopted:
$\bar{E} = k(x_1, x_2, \dots, x_n)$ and $\bar{f}(x) = (x - x_1)(x - x_2) \dots (x - x_n) = x^n + \bar{a}_1x^{n-1} + \dots + \bar{a}_n$ where the $\bar{a}_i$ are symmetric functions of the x_j. We saw that, under the n! obvious automorphisms of $\bar{E}$, the fixed field $\bar{F}$ was $\bar{F} = k(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_n)$ and $[\bar{E}:\bar{F}] = n!$ In Fig. A, we have the two situations $\bar{E}/\bar{F}$ and E/F, which we shall show to be isomorphic. Consider the mapping $\sigma: k[a_1, a_2, \dots, a_n] \longrightarrow k[\bar{a}_1, \bar{a}_2, \dots, \bar{a}_n]$ given by σ the identity on k and $\sigma(a_i) = \bar{a}_i$ where, for instance, $k[a_1, a_2, \dots, a_n]$ consists of the polynomials in the n variables. Being a substitution mapping, it is a homomorphism, and obviously it is onto. To show that it is an isomorphism, we need only that

$$\begin{tikzcd} E \arrow[r, "\tau'"] \arrow[d, dash] & \bar{E} \arrow[d, dash] \\ F \arrow[r, "\tau"'] \arrow[from=2-1, to=1-2, phantom, "\uparrow" description] & \bar{F} \end{tikzcd}$$

Fig. A

the kernel of σ is 0. Let $g(a_1, a_2, \ldots, a_n)$ be in the kernel. Then $g(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n) = 0$. This means, in terms of the independent variables $x_1, x_2, \ldots, x_n$, that $g\Big(\bar{a}_1(x_1, x_2, \ldots, x_n), \bar{a}_2(x_1, x_2, \ldots, x_n), \ldots, \bar{a}_n(x_1, x_2, \ldots, x_n)\Big) = 0$. As any substitution map is a homomorphism and will therefore yield 0 again, the substitution $x_i \longrightarrow \xi_i$ yields $g\Big(\bar{a}_1(\xi_1, \xi_2, \ldots, \xi_n), \bar{a}_2(\xi_1, \xi_2, \ldots, \xi_n), \ldots, \bar{a}_n(\xi_1, \xi_2, \ldots, \xi_n) = 0$. Sinde each $\bar{a}_i(\xi_1, \xi_2, \ldots, \xi_n)$ is the original a_i, we have $g(a_1, a_2, \ldots, a_n) = 0$ and σ is an isomorphism. σ can be extended to the quotient fields to obtain the isomorphism $\tau : F \longrightarrow \bar{F}$ such that $\tau(a_i) = \bar{a}_i$. Furthermore, this isomorphism of the ground fields can be extended to an isomorphism τ' of E and $\bar{E}$, the splitting fields of $f(x)$ and its image $\bar{f}(x)$, respectively. Thus, the general equation of degree n has the symmetric group S_n as Galois group. The proof that S_n, for $n > 4$, is not a solvable group may be found in any textbook on modern algebra.

5.7 Permutation Groups. We shall use the numbers 1, 2, 3, ..., n to represent the elements in a finite set. A one-one mapping σ of a finite set onto itself is called a permutation. A particular permutation can be described by giving $\sigma(i)$ for each $i = 1, 2, \ldots, n$. The cycle representation for σ is formed by selecting an element i and forming $\ldots, \sigma^{-1}(i), i, \sigma(i), \sigma^2(i), \ldots$ until repetitions occur and, then, repeating this process, for an i not in a previous cycle. If, for example, $\sigma(i) = 5, 4, 2, 3, 1$ for $i = 1, 2, 3, 4, 5$ respectively $\sigma = (1, 5)(4,3,2)$. If an element does not occur, it is assumed that it is left fixed. The cycle representation of σ is obviously unique. The 2-cycles (i, j) are called transpositions. Cycles will be compounded from right to left. It is easy to verify the following:

1) $(r, 1)(r - 1, 1) \ldots (3, 1)(2, 1) = (1, 2, 3, \ldots, r)$,

2) $(1, 3)(1, 2) = (1, 2, 3)$,

3) $(1, 2)(1, 3) = (1, 3, 2) = (1, 2, 3)^{-1}$,

4) $(1, 2)(1, 3)(1, 2) = (2, 3)$.

By 1) every cycle can be written as a product of transpositions. Since every permutation can be broken into cycles, it follows that every permutation can be written as a product of transpositions, which may be assumed to have the form $(1, i)$ by 4). Consider the field $Q(x_1, x_2, \ldots, x_n)$ and let $P = \prod_{i<j} (x_i - x_j)$. Let $\sigma(P)$ be the result of applying the permutation σ to the subscripts defining P. Then $\sigma(P) = (\text{sign } \sigma) \cdot P$ where sign σ is $+1$ or -1. The mapping given by $\sigma \longrightarrow \text{sign } \sigma$ is a homomorphism of the group S_n of all permutations of n elements onto $\{+1, -1\}$ (the transposition $(1, 2)$ has sign -1, as is easily shown) : $(\text{sign } \sigma\tau) \cdot P = \sigma\tau(P) = \sigma(\text{sign } \tau \cdot P) = \text{sign } \tau \cdot \text{sign } \sigma \cdot P$. The kernel of this homomorphism is called the alternating group A_n. Therefore, A_n consists of those σ such that $\sigma(P) = P$, and $S_n/A_n \simeq \{+1, -1\}$. A_n does not have invariant subgroups, except for $n = 4$, and $1 \subset A_n \subset S_n$ is the composition series. As $\frac{1}{2}(3!)$ is the prime 3, S_3 is solvable. S_4 can also be shown to be solvable. If $n > 4$, $\frac{1}{2}(n!)$ is not prime and S_n is not solvable.

Let σ be a permutation and $\sigma(i) = j$, i.e. i, j are neighbors in a cycle of σ. In terms of the cycles of σ, what can be said about the cycles of $\pi\sigma\pi^{-1}$ where π is another permutation? Contention: $\pi(i)$, $\pi(j)$ are neighbors in a cycle of $\pi\sigma\pi^{-1}$. This is evident since $\pi\sigma\pi^{-1}\big(\pi(i)\big) = \pi\sigma(i) = \pi(j)$. From $(1, 2) \notin A_n$ it follows that no transposition belongs to A_n for, if π is a permutation with $\pi(1) = i$ and $\pi(2) = j$ $(i \neq j)$, then $\pi \cdot (1, 2) \cdot \pi^{-1} = (i, j)$. By expressing a permutation

in terms of transpositions, one can decide whether or not it belongs to A_n according as the number of transpositions is even or odd. For $\sigma \varepsilon A_n$, we can even write $\sigma = (1, a)(1, b)(1, c)(1, d) \ldots$ where the number of transpositions is even. These transpositions may be grouped into successive pairs, and between the elements of each pair $(1, 2)(1, 2)$ may be inserted giving, e.g., $(1, a)(1, b) = (1, a)(1, 2)(1, 2)(1, b) =$ $(1, 2, a)(1, b, 2) = (1, 2, a)(1, 2, b)^2$. Consequently, $(1, 2, 3)$, $(1, 2, 4)$, ..., $(1, 2, n)$ are generators of A_n.

Any subgroup of the symmetric group is called a permutation group. A permutation group is said to be transitive if it changes every digit into every other digit. To show that a permutation group is transitive it is clearly sufficient to show that the group moves 1 into every digit.

Theorem: Let p be a prime, and let G be a transitive group with p letters such that G contains a transposition, say $(1, 2)$. Contention: G is a symmetric group, i.e. $G = S_p$.

Proof: Suppose G contains precisely the following transpositions containing 1: $(1, 2)$, $(1, 3)$, ..., $(1, r)$ where $2 \leq r \leq p$. We shall show that $r|p$ and, consequently, $r = p$, which implies that G contains all transpositions $(1, j)$. Since every permutation can be written in terms of these, it will follow that $G = S_p$. We have only to show that $r|p$. Apply one of the permutations σ of G to the set $S = \{1, 2, \ldots, r\}$. We shall show that either $\sigma(S)$ and S have no digit in common or $\sigma(S) = S$. Suppose $i \varepsilon S \cap \sigma(S)$ and $r \div 1 \varepsilon \sigma(S)$ where, say, $i = \sigma(a)$ and $r + 1 = \sigma(b)$. G, containing the symmetric group S_r of the first r digits, also contains (a, b). Consequently, by group properties $\sigma \cdot (a, b) \cdot \sigma^{-1} = (i, r + 1)$, and G contains $(i, 1)$, $(i, 2)$, ..., $(i, i - 1)$, $(i, i + 1)$, ..., (i, r), $(i, r + 1)$. Hence, G contains S_{r+1} and, in particular, $(1, r + 1)$, which

is a contradiction. Therefore, $S \cap \sigma(S) = \emptyset$ or S. For $\sigma, \tau \in G$ we can, then, say $S \cap \sigma^{-1}\tau(S) = \emptyset$ or S, which yields upon application of σ that $\sigma(S) \cap \tau(S) = \emptyset$ or $\sigma(S)$. Considering S and all possible images of S, we can say that any two have nothing in common or coincide. However, the transivity of G gives $\{1, 2, \dots, p\} = \bigcup_{\sigma \in G} \sigma(S)$. Thus, $\{1, 2, \dots, p\}$ can be written as the union of disjoint sets, each containing r elements, and, therefore, $r | p$.

Exercise: If in the theorem "transposition" is replaced by "3-cycle," show that G is either the symmetric group or the alternating group.

For an application of theorem A to Galois theory, consider: F, a ground field; f(x), separable of degree n, $\in F[x]$; E, the splitting field of f(x), where $f(x) = (x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n)$ and $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. As $\sigma \in G$ is described by knowing all $\sigma(\alpha_i)$, f(x) gives G the additional structure, viz. G is a permutation group. Note, however, that this permutation group is not an invariant of only E/F. Further, f is irreducible $\Longleftrightarrow$ G is transitive: since f irreducible $\Longrightarrow$ there exists a σ such that $\sigma(\alpha_1) = \alpha_i$ for each i, and f reducible $\Longrightarrow$ α_1 is moved only into the roots of the irreducible factor and G is not transitive. Consider the situation in Fig. A where Q is the field of rational numbers, R the reals, C the complex numbers, and E is the splitting field of $f(x) = x^5 - 10x + 2$. f(x) is irreducible by Eisenstein's criterion, and it is easy to show that it has exactly two complex roots. The Galois group G is transitive and C = ER. We also know that H consists of the

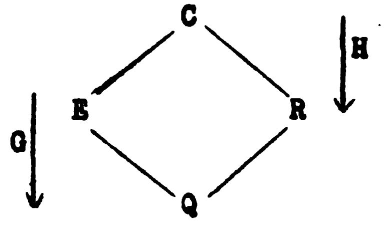


Fig. A

identity and conjugation, which permutes the two complex roots and leaves the others fixed. By the Lagrange theorem of natural irrationalities, it follows that, say, $(1, 2) \varepsilon G$. Thus, f(x) has S_5 as Galois group and is not solvable. In this example f(x) can be generalized by requiring only that it be irreducible of prime degree with exactly two complex roots.